Color Evolution of Kaede-type Red Fluorescent Proteins

by

Hanseong Kim

A Dissertation Presented in Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

Approved April 2012 by the
Graduate Supervisory Committee:

Rebekka M. Wachter, Chair
Petra Fromme
Kevin E. Redding

ARIZONA STATE UNIVERSITY

May 2012

ABSTRACT

The green fluorescent protein (GFP)-like fluorescent proteins play an important role for the color of reef-building corals. Different colors of extant coral fluorescent proteins (FPs) have evolved from a green ancestral protein. Interestingly, green-to-red photoconversion FPs (Kaede-type Red FPs) are only found in clade D from Scleractinia (Faviina suborder). Therefore, I focus on the evolution of Kaede-type FPs from Faviina suborder ancestral FP. A total of 13 mutations have been identified previously that recapitulate the evolution of Kaede-type red FPs from the ancestral green FP. To examine the effect of each mutation, total ten reconstructed FPs were analyzed and six x-ray crystal structures were solved. These substitutions created a more hydrophilic environment around the carbonyl group of Phe61. Also, they increased the flexibility of the c-terminal chain, which keeps it from interacting with the entrance of the putative solvent channel. The photoconversion reaction shows a two-phase kinetics. After the rapid initial phase, the overall reaction followed the first-order kinetics. Based on the crystal structure analysis, I propose a new mechanism for Kaede-type FP photoconversion process, which a proton transfers via Gln38 to the carbonyl group of Phe61.

DEDICATION

To My Mother, Father and Brother.

ACKNOWLEDGMENTS

This Dissertation would never have been written without the guidance and the help of several individuals who in one way or another contributed and extended their valuable assistance in the preparation and completion of this study.

First and foremost, I would like to express my sincere appreciation and gratitude to my graduate advisor, Dr. Rebekka M. Wachter. She has been an incredible mentor and her excellent advice guides me to. She supports me no matter where I am and what I do. I always feel lucky to work with her.

I also would like to thank Dr. Petra Fromme and Dr. Kevin E. Redding for their helpful comments and discussion on my dissertation.

I want to thank past and present members of Wachter lab; Dr. J. Nathan Henderson, Dr. Jennifer Watkins, Dr. Di Wu, Gabi Malo, Alisson Chausse, Alan Conrad, Suratna Hazra, Agnieszka Kuriata, Dayna Peterson, and Tim Grunkemeyer. They always gave me the good advice and friendship all these years in the lab.

I want to thank Dr. Raimund Fromme and Dr. Liqing Chen for helping me to learn about X-ray crystallography.

I also want to thank Dr.Mikhail V. Matz and Chintan Modi for the collaboration.

Finally, I'm very grateful to my parent. Although I have been a really bad son to them, they have been an endless source of love and support and have always encourage me to strive to attain my goal. I really love you.

TABLE OF CONTENTS

Page

Page

LIST OF TABLES

LIST OF FIGURES

Figure Page

LIST OF SCHEMES

INTRODUCTION

Most naturally occurring green fluorescent protein (GFP)-like proteins are highly fluorescent when exposed to ordinary day light, emitting electromagnetic radiation within distinct color classes such as cyan, green and red (*1, 2*). The ability of fluorescent proteins (FPs) to emit visible light stems from a post-translational self-modification of three consecutive amino acid residues that adopt a distorted helical conformation in the interior of the native protein (*3*). When this tripeptide is subjected to a series of autocatalytic steps involving O_2-dependent and O_2-independent processes, a light-responsive chromophore is generated (Figure 1). The extent of bond desaturation within the chromophore coupled with the chemical environment of the surrounding protein matrix dictates the wavelength of light emitted by the protein. Manipulation of these parameters has allowed for color tuning within the visible and near-visible range of electromagnetic radiation and a broad variety of engineered FPs are currently available as research tools in biotechnology (*4, 5*).

GFP-like proteins consists of an 11 stranded β-barrel and a α-helix which contains the chromophore. The transition from the natively folded precursor protein to the green fluorescing mature form is governed by three major chemical steps. In avGFP, which is derived from the Pacific Northwest jellyfish *Aequoria victoria*, the functional tripeptide consists of a Ser-Tyr-Gly sequence in positions 65-67. Although only the second and third residues are conserved, the tripeptide is always located within a helical segment that traverses the center of the protein (Figure 1). The protein fold encourages nucleophilic attack of the Gly67 amide

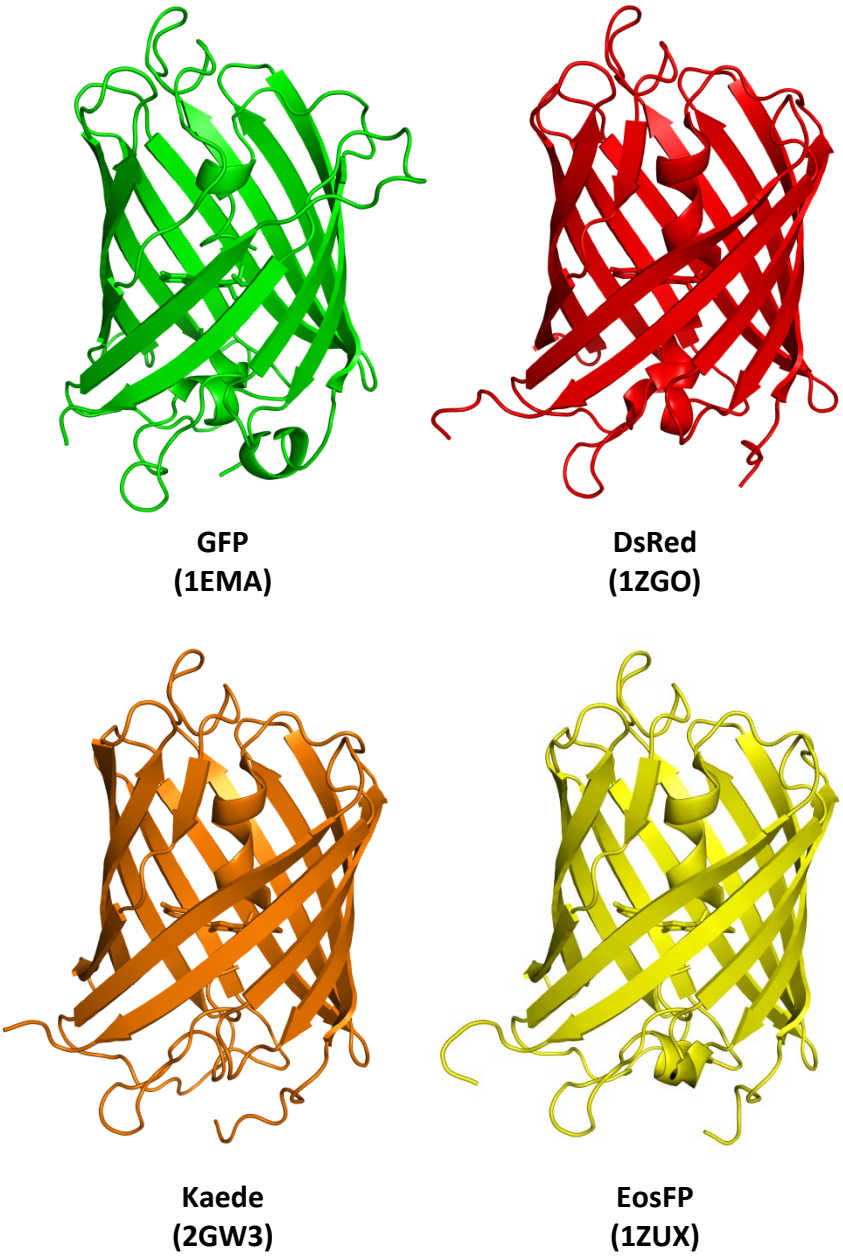
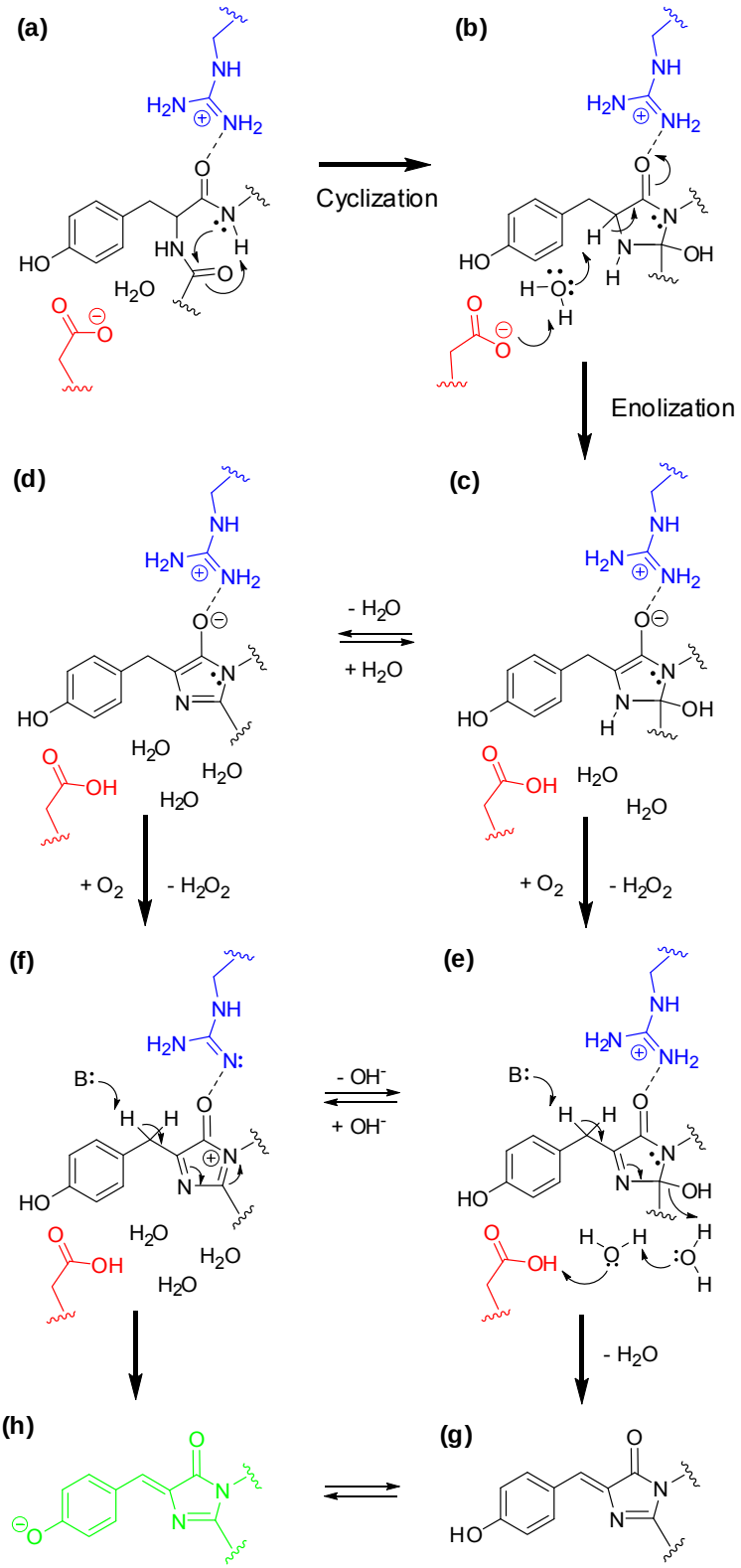


Figure 1. Ribbon diagram of crystal structure of GFP-like proteins. Chromophores are shown in stick representation.

Scheme 1. Mechanism of Chromophore Formation in Green Fluorescent Protein (*3*, *6*, *7*)



nitrogen on the carbonyl carbon of Ser65 (species a in Scheme 1), creating a heterocyclic structure from the backbone atoms of the three residues (species b in Scheme 1). The cyclization reaction is promoted by the electrostatic catalyst Arg96, and subsequent removal of the Tyr66 α-proton is likely facilitated by the acid-base catalyst Glu222 (*8, 9*). The resulting α-enolate form of the five-member ring (species c and d in Scheme 1) is thought to be reactive to two-electron oxidation by molecular oxygen, generating the cyclic imine form of the protein and hydrogen peroxide as a side product (species e and f in Scheme 1) (*10*). Several years ago, the cyclic imine intermediate was trapped by mutagenesis, and its crystal structure was determined in the hydrated state (e), as well as in several partially dehydrated states (*11, 12*). Subsequently, deuterium kinetic isotope effect data provided evidence that the cyclic imine undergoes deprotonation at the β-carbon of Tyr66 in a net water elimination process, thereby completing all biosynthetic steps necessary for generating the GFP-like chromophore (species g and h in Scheme 1) (*6*). Although it was long thought that the side chain of Tyr66 mediates the transfer of electrons to molecular oxygen, the kinetic results have repudiated this notion (*10*). Instead, the phenolic group appears to play an essential role in the final dehydration step that contributes to overall rate retardation. Therefore, GFP maturation is best described by a cyclization-oxidation-dehydration sequence of events, where the oxidation step is solely guided by main chain chemistry (*11, 12*). The mature GFP-like chromophore typically emits bright green light in its anionic form (B-form, species h in Scheme 1) and is usually non-fluorescent in its neutral form (A-form, species g in Scheme 1). However, upon excitation of the A-form, green fluorescence may be acquired by excited-state proton transfer (ESPT).

The GFP folding topology consists of a barrel-shaped domain, in which 11 β-strands assemble to enclose a helical structure that may or may not bear a chromophore (Figure 1). Aside from GFP-like proteins, this fold has been identified in human extracellular matrix proteins. In these proteins, the 11-strand β-barrel termed G2F (globular-2) fragment is part of a multidomain protein and provides a binding surface for protein - protein interactions (*13*). Therefore, two major gene lineages harbor proteins adopting the GFP fold within the kingdom animalia, the colorless G2F lineage and the GFP-like lineage. These two branches of the protein superfamily tree are thought to have originated from a common ancestor that predates the divergence of Radiata (phylum Cnidaria) and Bilatera (phyla *Arthropoda, Chordata*, etc.) (*14*).

Since the discovery of GFP from the jellyfish *Aequorea victoria*, a large number of GFP-like proteins have been identified in class *Anthozoa*. Although the biological purpose of FPs is still unknown, it is believed that coral FPs are a major contributor of the color of corals (*15-17*). Previous research has shown that there are four basic color groups in coral FPs; green, cyan, red and non-fluorescent purple-blue. Within class *Anthozoa*, four ancient FP clades have been classified and exhibit extensive color diversification (Figure 2) (*18, 19*). Clade A has been assigned to order *Actiniaria* comprising the sea anemones, and clades B-D have been assigned to order *Scleractinia* comprising mostly the reef-building or stony corals. Clade B constitutes a lineage with functional specialization related to the purple phenotype, essentially nonfluorescent chromoproteins as well as red FPs. Clade C consists of three subclades (C1, C2 & C3) and includes cyan, green, yellow and red FPs. Clade D FPs falls into cyan,

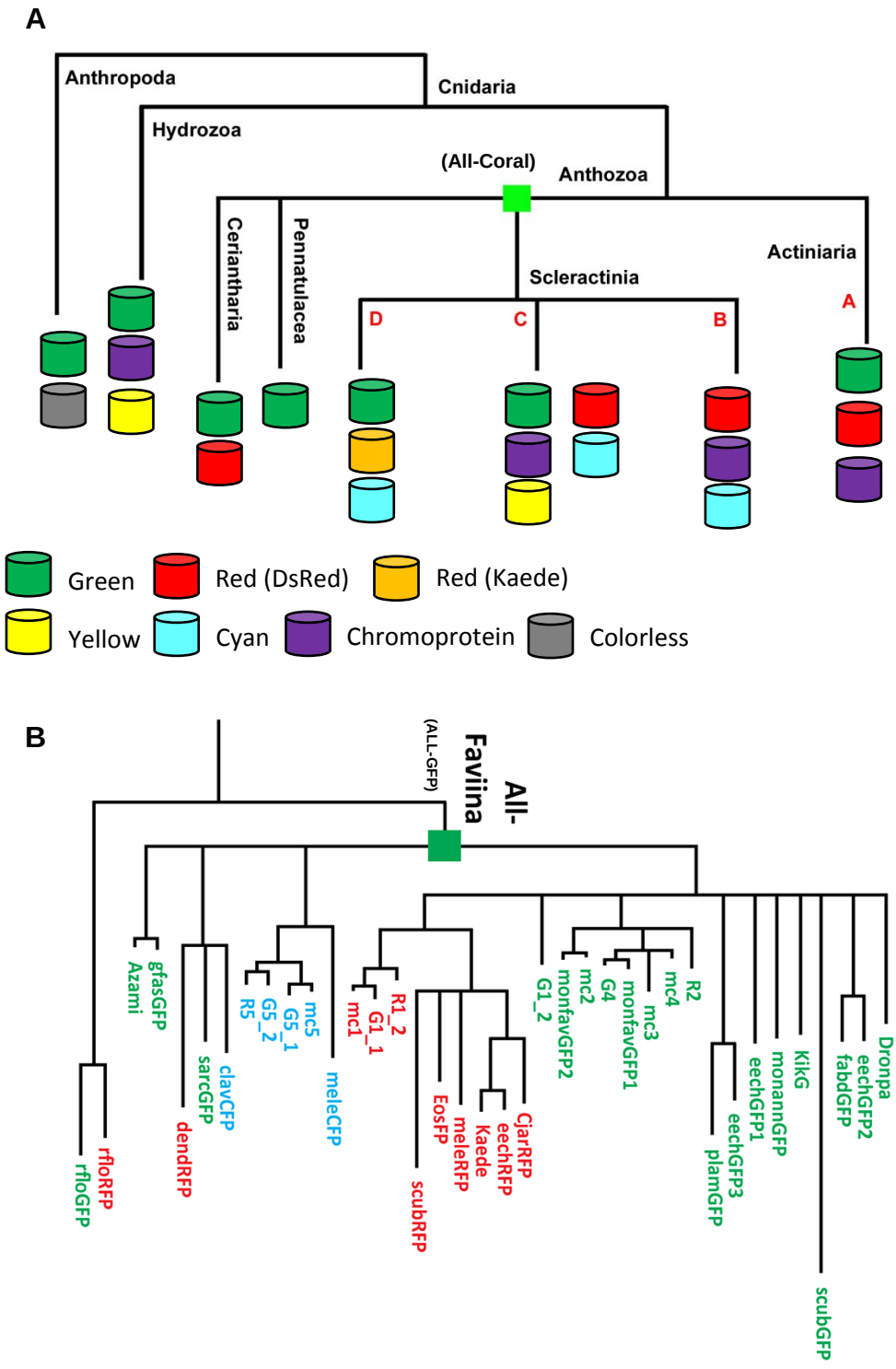


Figure 2. Phylogenetic trees of GFP-like proteins (*19*) - (A) Outline of major phylogenetic branches of the GFP-like proteins. (B) The phylogenetic tree of clade D fluorescent proteins.

green and Red FPs (*19*). Regardless, the common ancestor to all coral fluorescent proteins is thought to be of the green phenotype (*20*), whereas color diversification appears to be a more recent event that occurred upon positive selection for specific color classes in response to environmental pressures (*21*). Although over the years a variety of evolutionary driving forces have been suggested, the biological purpose and necessity for specific color classes remains poorly understood (*21*). Over the past decade, extensive protein engineering efforts have provided an impressive array of red and far-red fluorescent proteins useful in deep tissue imaging of mammals (*22*), useful in the live cell observation of single-molecule dynamics (*23*), and useful in the provision of photoactivatable tools for super-resolution microscopy (*24*).

Photoactivatable Fluorescent Proteins

In addition to a wide range of colors of GFP-like proteins, a number of different variants of FPs, which are called molecular highlighters, optical highlighters, or photoactivatable fluorescent proteins, have been found. These variants exhibit some interesting properties including photoactivation, photoswitching, and photoconversion. These unusual optical properties play an important role in recent advances of biological imaging and analysis (*25-29*). Photoactivatable fluorescent proteins divide into two groups. The first group of FPs (photoswitching FP) reversibly changes between a fluorescent and a non-fluorescent state by irradiation (*30-32*). Among the photoswitching FPs Dronpa (*30*), asFP595 (*33*), and mTFP0.7 (*31*) are the most well-known FPs. Dronpa converts to a non-fluorescent state with 488 nm light, and switches back to a

fluorescent state with 405 nm light (*30*). A similar mechanism is suggested for mTFP0.7 and KFP1. The second group of FPs (photoactivation and photoconversion FPs) irreversibly converts between two different fluorescent states. Photoactivatable FPs transform from a non-fluorescent to fluorescent state by irradiation (*34-36*). For PA-GFP (*34*) and PS-CFP (*36*), the anionic chromophore can be generated from irreversible decarboxylation of Glu222 by irradiation of UV light. Photoconversion FPs can change their color from green to red with UV or violet light (*37-40*). These 'Kaede-type' FPs, which include the His-Tyr-Gly chromophore, generate the red chromophore via an unusual peptide backbone cleavage reaction (*39, 41-43*).

Green-to-red (G/R) photoconversion proteins: Kaede-type red fluorescent proteins.

Discovery of G/R Photoconversion: Kaede, EosFP, DendFP, and Relatives.

Three years after the discovery of DsRed, the first discovered red fluorescent protein, a unique type of red fluorescence was reported with maturation features entirely distinct from those of DsRed (*41*). In 2002, Miyawaki and co-workers reported the astounding observation that the green-fluorescing form of a newly cloned GFP-like protein was modified to a bright red species when exposed to sunlight. This protein, derived from the stony coral *Trachyphyllia geoffroyi*, was termed Kaede (maple leaf in Japanese), based on the observation that G/R color conversion appeared to be triggered by UV or violet light (*41*). The irreversible red shift in emission from 518 nm (green) to 582 nm (red) occurred with a low quantum yield of 2.4×10^{-4} upon 400 nm light exposure (*41*). Shortly afterward,

Nienhaus and co-workers described a similar photoconversion process in a fluorescent protein termed EosFP from the stony coral *Lobophyllia hemprichii*, a protein with an amino acid sequence 84% identical to that of Kaede (*39*).

Subsequently, another close relative termed DendFP from *Dendronephthya* (*18, 19*) was shown to share the same phenomenon of light-induced color modification (*44*). In 2006, an engineered monomeric version, Dendra2, was described to efficiently photoconvert when irradiated with longer-wavelength light such as 488 nm light (*40*). In the same year, the fluorescent proteins mcavRFP and rfloRFP, cloned in 2002 from the great star coral *Monastrea cavernosa* and the coral limorph *Ricordea florida* (*18, 19*), respectively, were shown to acquire red color as a function of light stimulation (*14*). In combination, the experimental evidence strongly suggested that these kinds of proteins are widespread within clade D of the stony corals (Figure 2).

Light-Induced Main Chain Scission Is Intrinsic to G/R Photoconversion and requires a Histidine Residue.

All known members of the Kaede-type class of red fluorescent proteins bear a His-Tyr-Gly tripeptide at positions 62-64 (Kaede). The histidine residue is critical for the formation of the red chromophore, because its imidazole ring is ultimately incorporated into the three-ring π-system of the mature chromophore (Scheme 2). Support for this notion was provided early by the observation that G/R photoconversion was abolished upon substitution of His65(62) with phenylalanine, glutamine (*41*), or any other amino acid residue (*37*). In the same

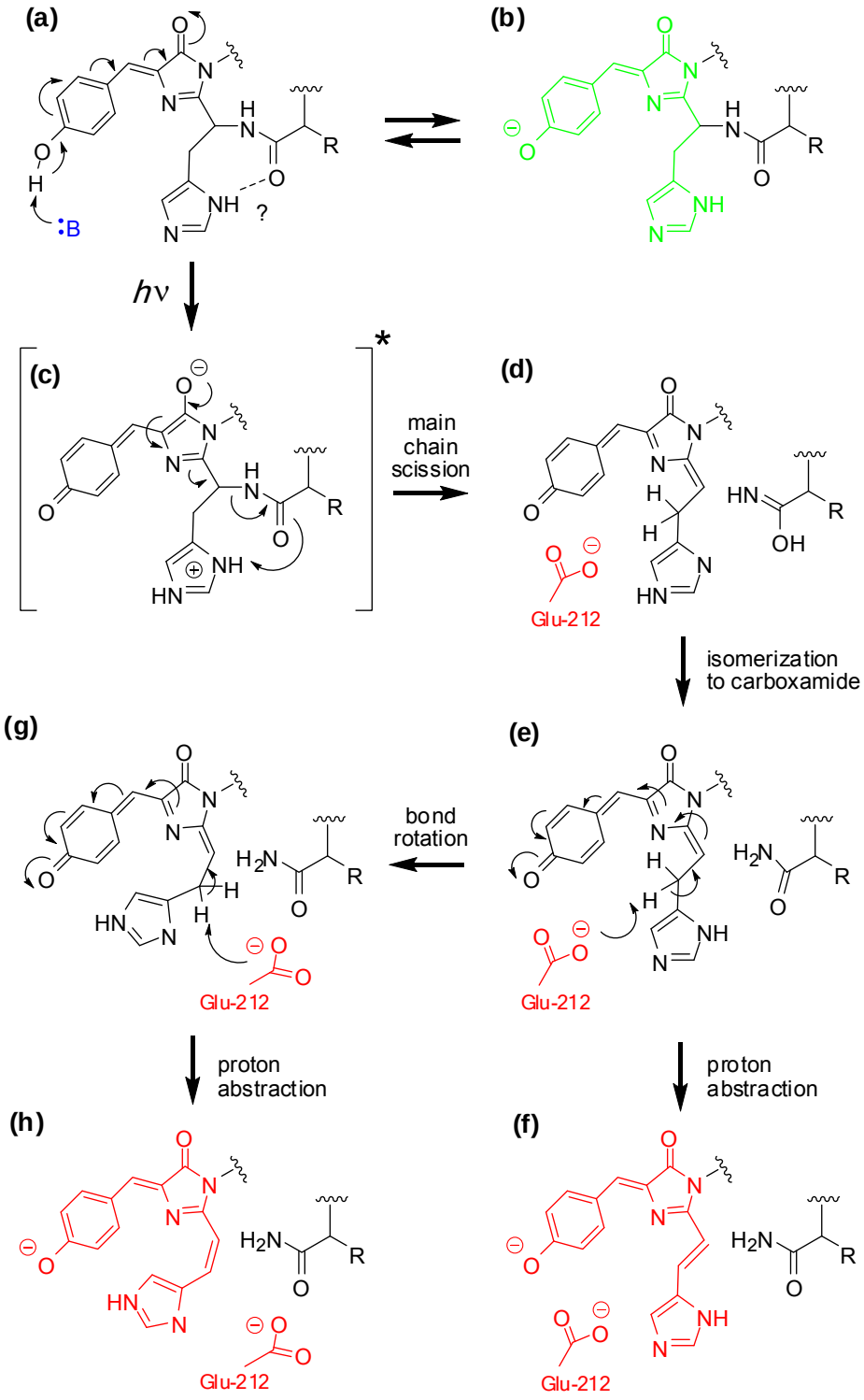
study, the chemical structure of the red chromophore was elucidated via application of NMR methods to a five-residue tryptic peptide bearing the chromophore (*37*). The π-conjugated system was shown to include the His65(62) side chain, based on a desaturated bond between C_α and C_β of His65(62) (Scheme 2A and B) (*37*). The authors demonstrated further that the A69S substitution in the chromophore environment abolished the photoconversion chemistry, although green color was retained (*37*). In the absence of an appropriate light source for photochemical modification, the green form of Kaede is the final state of protein self-processing. This form bears a *cis* coplanar two-ring chromophore chemically identical to that found in GFP, likely synthesized in a similar manner and attached to a single polypeptide chain. As originally described for GFP variants (*45-47*), the GFP-like chromophore of Kaede is pH-titratable in its ground state, with the absorbance band at 380 nm arising from the neutral (phenolic) form and the band at 508 nm arising from the anionic (phenolate) form (Scheme 1) (*41*). In 2003, the extraordinary discovery was made that the light-induced chemical changes are coupled to an unusual main chain cleavage reaction occurring in the interior of the β-barrel, without disruption of the tertiary structure (*37*). A careful mass spectrometric analysis of tryptic peptides derived from green and red forms of Kaede led to the identification of His65(62) as the site for backbone scission, generating two protein chains with respective masses of 10 and 18 kDa. The break in main chain connectivity was inferred to result from an elimination reaction, in which the His65(62) amide nitrogen is ejected from the adjacent α-carbon (Scheme 2). In addition, the mass spectral results were in line with the formation of a carboxamide group at the

newly formed C-terminus of the peptide fragment comprising the N-terminal 10 kDa segment of the precursor chain (Scheme 2) (*37*).

Photoconversion Occurs only from the Nonfluorescent Form of the GFP-like Chromophore.

In their seminal work, Miyawaki and co-workers proposed that electronic excitation of the neutral form of the GFP-like chromophore triggers excited state deprotonation of its phenolic hydroxyl group (species a in Scheme 2), consistent with ESPT as first described for GFP some years ago (*48*). This process was suggested to produce a transient excited state I* consisting of the chromophore anion, a state that may subsequently decay by β-elimination (conversion of species c to species d in Scheme 2) (*37*). According to this idea, light activated ejection of the amide leaving group from the His65(62) α-carbon would yield a hydroxyimino group, a process that would be facilitated by protonation of the Phe64(61) carbonyl oxygen. This event would be followed by rapid isomerization of the hydroxyimino group to the carboxamide form (species e in Scheme 2) (*37*). Full maturation to the red-emitting form would require an additional step entailing proton abstraction from the β-carbon of His65(62), such that the ensuing C_α-C_β double bond provides for conjugation of the histidine imidazole ring with the remainder of the chromophore skeleton (species f in Scheme 2). The result is a red-emitting three-ring chromophore that experiences reversible protonation-deprotonation equilibria similar to those observed for the green form. The pKa value of the red chromophore of Kaede was reported to be 5.6 (*41*) and 6.0 (*49*) in separate studies, and as in the green species, steady-state fluorescence

Scheme 2.Hypothetical Mechanism of Red Chromophore Formation in Kaede-Type G/R Photoconversion Proteins (*7*).



emission occurs exclusively from the anionic state of the red chromophore (*49*). In 2004, clear evidence was provided that only the neutral form of the green state is competent in undergoing light-triggered modification to the red form (*39*) . In this work, it was demonstrated that the photochemical action spectrum of EosFP follows the absorption intensity of the GFP-like neutral chromophore almost exactly, consistent with a mechanism in which removal of an excited-state proton from the chromophore's phenolic end may play an essential role (*39*). According to this model, a proton would be transferred from the phenolic side of the chromophore to the imidazole ring of His65(62) through the protein matrix over a distance of more than 10 Å. This process has been proposed to yield a transient imidazolium cation (species c in Scheme 2) that may serve as proton donor to facilitate the β-elimination reaction (*39*).

Long-range ESPT in G/R photoconversion remains highly controversial.

A strong argument against ESPT is the complete lack of green fluorescence upon excitation of the neutral chromophore (*50*). Importantly, a hydrogen-bonded network that would facilitate the transfer of a proton from the chromophore to His65(62) has not been identified in any of the crystal structures (*50*). To date, ESPT has been demonstrated solely in the red version of Kaede, where the excitation of the protonated three-ring chromophore results in the emission of red light from the anion (*49*). A comparison of crystal structures of the green and red states of EosFP (*50*) and Kaede (*51*) suggests that the H-bonding pattern around the phenolic end remains largely unaffected by photoconversion (Figure 3A), implying that potential proton transfer pathways may be similar in the green

and red forms of these proteins. However, no experimental information is available with respect to the nature of the proton acceptor, the long-term residence of the proton upon dissociation from the chromophore, or any solvent equilibration effects. The pKa value of green EosFP was shown to be 5.8, very close to that of Kaede, and in line with the higher photoconversion efficiency observed at lower pH values (*39*). Surprisingly, green Dendra2 was shown to exhibit a pKa value of 7.1, a significantly elevated proton dissociation constant compared to those of EosFP and Kaede, likely a consequence of conformational adjustments of Arg69(66) (*52*). The pKa value for the red chromophore of Dendra2 was found to be similarly elevated to 7.5. These values are noteworthy, because they imply that in a solution of pH <7, photoconversion of Dendra2 occurs with a protonated chromophore as both the reactant and product (*52*). This means that a proton released in the excited state would have to rapidly reattach itself to establish the appropriate product equilibrium, eliminating the primary rationale for invoking long-range proton transfer. In further support of a model that lacks ESPT as a mechanistic feature, a recent theoretical study has suggested that the chromophore's phenolic end remains protonated throughout the photochemical reaction progress (*53*).

Structural Preorganization and Glu222(212) as a General Base for Proton Abstraction

When X-ray structures of green and red forms of EosFP were compared, the unmodified His65(62) and its oxidized equivalent were found to adopt an almost identical extended side chain conformation (Figure 3) (*50*). This feature was

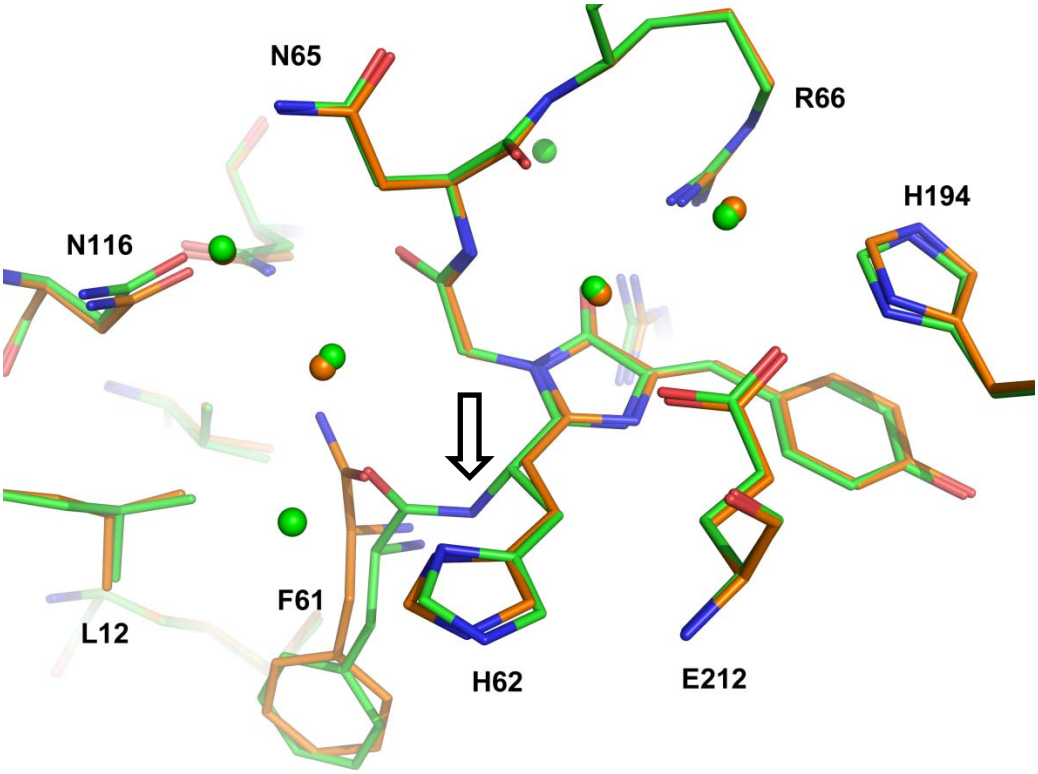


Figure 3. Superimposition of EosFP green (PDB entry 1ZUX colored green) and red forms (PDB entry 2BTJ colored orange). The site of bond scission upon G/R photoconversion is indicated by an arrow.

subsequently confirmed in Kaede (*51*), and structural adjustments due to green-to-red conversion were therefore judged to be minimal in both proteins. In EosFP, a slight rotation of the imidazole group was observed, bringing the ring into the plane of the remainder of the chromophore (Figure 3), whereas in the green form of Kaede, the imidazole ring was found to be perfectly coplanar with the GFP-like chromophore, suggesting a large degree of structural preorganization. In both proteins, the carboxyl of Glu222(212) was found to be positioned near the β-carbon of His65(62) at a distance of 3.3Å, poised perfectly for general base catalysis in abstraction of a proton from the His62 C_β atom (Scheme 2) (*50*). Convincing support for the function of Glu222(212) as a general base catalyst was provided by the observation that its replacement with a glutamine abolished red fluorescence entirely (*50*), as was subsequently confirmed with Kaede (*51*).

A few years ago, Nienhaus and co-workers proposed a one-step elimination mechanism, primarily based on the lack of major structural adjustments upon photomodification (*50*). This mechanism is distinct from the previously proposed E1-type mechanism shown in Scheme 2 (*37*), as the deprotonation step at the His65(62) C_β atom would be responsible for triggering C-N bond cleavage, rather than follow it. An important aspect of such a concerted elimination mechanism is the fact that stereoelectronic control would be provided by the trans stereochemistry of the histidine side chain (*50*). However, recent experimental and theoretical work argues against a one-step deprotonation elimination mechanism (see below) (*42, 53*). So far, the trapping of intermediate forms of Kaede by mutagenesis has remained unsuccessful, in line with short-lived, unstable intermediate species (Scheme 2) (*51*).

Proposed Role of the His65(62) Imidazole Group as a Proton Donor

In the crystal structures of the green and red forms of EosFP (*50*), the His65(62) $N_{\delta 1}$ atom was found to be positioned within hydrogen bonding distance of the Phe64(61) carbonyl oxygen, whereas in Dendra2, the equivalent distance was increased to 3.6 Å (*52*). On the basis of this observation, Nienhaus and coworkers suggested several years ago that His65(62) may be the transient acceptor of a light-triggered proton transfer reaction from the phenolic hydroxyl of the chromophore to the imidazole ring (*50*), and that the imidazolium cation would subsequently donate a proton to the main chain carbonyl of Phe64(61), thus facilitating the elimination reaction (species c in Scheme 2) (*50*). Unfortunately, the hydrogen bonding geometry appears to be somewhat unfavorable, as the carbonyl lone electron pair would be oriented nearly perpendicular to the imidazolium N-H bond proposed to serve as a donor. Miyawaki and co-workers have noted that the green versions of Kaede and EosFP bear a positionally conserved and highly ordered water molecule adjacent to and in plane with the imidazole ring of His65(62) (Figure X) (*51*). Although the hydrogen bonding geometry is not optimal for this interaction either, the additional solvent molecule may aid in stabilizing a transiently charged imidazolium cation.

Recently, computer simulations have been used to shed light on the catalytic role of the imidazole group of His65(62) (*53*). Using hybrid quantum chemical and molecular mechanical potentials, the theoretical results appear most consistent with a neutral His65(62) imidazole ring serving as a proton donor to

the carbonyl of Phe64(61), such that a transient imidazolide anion would be generated during the singlet excited state of the chromophore. Subsequently, the system would undergo intersystem crossing to the triplet state, from which peptide bond cleavage would proceed. In the final step, the imidazolide anion would promote abstraction of a proton from C_β, whereas the carboxylate of Glu222(212) would play the role of an electrostatic catalyst (*53*). Although this mechanism is novel and noteworthy, as it details mechanistic features without invoking phenolic ESPT, the exact catalytic function of His65(62) as a proton transfer agent remains unresolved, and a passive role for Glu222(212) in proton abstraction seems unlikely.

Engineering of KikG for G/R Photoconversion Provides Support for a Two-Step Elimination Mechanism.

The fact that His65(62) participates in the G/R photoconversion machinery (*37*) was exploited in the engineering of a photoactive variant from a natively green protein termed KikG (*38*). First, Asp65(62) was replaced with a histidine residue, followed by more extensive redesign of the chromophore-bearing pocket that introduced a total of eight substitutions (*38*). The resulting clone, named KikGR, demonstrated highly efficient G/R photoconversion. Unlike Kaede and EosFP, the green chromophore of KikGR was shown to titrate with the surprisingly high pKa value of 7.8 (*38*). Therefore, the neutral form of the chromophore dominates at physiological pH, providing a rationale for the much improved efficiency of photoconversion. Similar to those of other red forms, the absorbance spectrum of the photo-modified chromophore exhibits two maxima at 360 and 583 nm, both

arising from the anion. Excitation of either leads to 593 nm emission with high quantum yields of 0.92 and 0.65, respectively (*38*). Screening of KikGR mutants for crystallization led to the successful crystallization of a clone termed KikGRX bearing four additional substitutions (*42*). Pre- and post-photoconversion structures were determined, and in its red form, the chromophore of KikGRX was shown to adopt a distinctly *cis* conformation along the desaturated C_α-C_β bond of His65(62) (species h in Scheme 2). This configuration is opposite to the usual trans-isomers found in all other G/R proteins (species f in Scheme 2). Therefore, the authors have proposed that the rearrangement in the red chromophore structure results from conformational adjustments of an intermediate form generated upon main chain scission. Furthermore, the authors have suggested that in the GFP-like precursor state (species a and b in Scheme 2), rotation around the His65(62) C_α-C_β bond is restricted due to H-bonding of the imidazole ring with the carbonyl oxygen of Phe64(61) (*42*). The crystal structures of KikGRX are particularly significant, because they include the first structure of a pre-photoconversion G/R protein, in which the chromophore is in the neutral state competent in photoconversion (species a in Scheme 2). In this state, the His65(62) side chain clearly adopts an extended conformation. Therefore, a likely explanation for the *cis* configuration of the red form is a rotation around the saturated C_α-C_β bond prior to abstraction of a proton from C_β. The existence of an intermediate with a lifetime that allows for side chain conformational changes provides strong support for a two-step E1-type elimination mechanism (Scheme 2). Once the proton has been removed from C_β, bond desaturation will provide for full conjugation of the imidazole ring with the remainder of the chromophore π-

system; therefore, the conformation will be locked once the protein has fully matured to the red form (species h and f in Scheme 2) (*42*).

Ancestral gene reconstruction

The last two decades have witnessed major advances in phylogenetics and DNA synthesis techniques. With help from molecular biology, biochemistry and structural biology, we are able to experimentally reconstruct ancestral genes, inferred from phylogenetic analysis, with the intention of functional testing (*54-56*). A study of ancestral proteins can reveal valuable information not easily available from extant proteins. This technique also can provide historical context of molecular evolution. A comparison between ancestral sequences and existing proteins has made it possible to identify when the mutations are introduced. In addition, protein structures of ancient genes can help to identify historical amino-acid replacements that are likely to be involved in functional evolution. This approach has been applied to several systems to explain the evolution of molecular structure and function, such as Angiotensin II (*57*), antiviral RNase (*58*), steroid receptors (*59-61*), and GFP-like proteins (*20, 62, 63*).

Generally, ancestral gene reconstruction consists of five different steps (*54*). First, sequences of extant proteins along with outgroup sequences are aligned and the phylogenetic tree is constructed. Second, ancestral sequences are inferred, based on the phylogeny that describes their historical relationships and a statistical model of amino acid substitution that describes the relative probability of replacing each amino acid with any other amino acid. Third, the

coding sequence is recreated by synthesizing overlapping oligonucleotides and assembling them by PCR or by restriction digestion and ligation. Fourth, the ancestral gene is cloned into a plasmid and the plasmid is then expressed in bacterial or mammalian cells. Finally, the ancestral protein is purified and its functions can be characterized using experimental tests. Structural studies of ancestral proteins can also provide insight to the mechanism of functional shifts between the ancestral and extant proteins.

Ancestral gene reconstruction of GFP-like proteins

The green and cyan FPs require two consecutive autocatalytic reactions to generate their chromophores, whereas red, purple and yellow fluorescent proteins need an additional modification step to generate extended π-system to their chromophores. Since GFP-like proteins can be easily expressed and screened based on their colors, they have been considered as a convenient model for basic studies in an evolution process at molecular level (*18-21, 64*). With the ancestral gene reconstruction technique, Matz and co-workers resurrected an ancestral GFP-like protein of suborder *Faviina* (ALL-GFP), which displayed green emission (*20*). In 2008, two ancestral proteins were also reconstructed, which provide perspective to the history of coral color evolution. One was all-coral ancestor, which is the common ancestor of all coral GFP-like proteins. The other was all-Faviina, an ancestor of all *Faviina* proteins, and this protein is an updated version of the ALL-GFP. Both all-coral ancestor and all-Faviina ancestor turned out to be green fluorescent proteins. The other colors

evolved from green through small transitions, starting as early as late Triassic – early Jurassic (*19*).

Matz and colleagues tried to identify the amino acid residues that are responsible for the evolution of Kaede-type red fluorescent protein in the great star coral *Montastrea cavernosa* (*62*). When comparing the sequences of extant red-fluorescent proteins of *M. cavernosa* with the set of sequences consistent with the ALL ancestral node thought to be green, the least divergent pair involves 37 amino acid substitutions. Recently, more sophisticated transitional libraries were generated by using this protein pair as a basis, in which each of the 37 sites was given roughly half of the sites in the ancestral residue and a half in the extant residue (*62*). Unexpectedly, partially evolved yellow-orange clones were only rarely observed, suggesting extensive epistatic interactions of a set of key residues. Statistical analysis allowed for the identification of 13 sites, including one deletion, as essential in the evolutionary transition from the green to the Kaede-type red phenotype. Among the set of 13 substitutions, only five of them were around the chromophore. Interestingly, the remaining eight were pointing outwards towards the solvent or residues more distant from the chromophore. These eight residues would have been difficult to identify using the structural observation and mutation study of extant proteins.

```
GFP             1  MSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYGKLTLK   45
ALLGFP_OLD_     1  MS----VIKSDMKIKLRMEGTVNGHKFVIEGEGEGKPYEGTQTMN   41
ALLGFP_NEW_     1  MS----VIKSDMKIKLRMEGTVNGHKFVIEGEGEGKPYEGTQTMN   41
ALLQ65H         1  MS----VIKSDMKIKLRMEGTVNGHKFVIEGEGEGKPYEGTQTMN   41
LEAX6           1  MS----VIKSDMKIKLRMEGTVNGHKFVIEGEGEGKPYEGTQTMN   41
LEA             1  MS----VIKSDMKIKLRMEGTVNGHKFVIEGEGEGKPYEGTQTMN   41
LEAX72          1  MS----VIKSDMKIKLRMEGTVNGHKFVIEGEGEGKPYEGTQTMN   41
LEAX121         1  MS----VIKSDMKIKLRMEGTVNGHKFVIEGEGEGKPYEGTQTMN   41
Kaede           1  MS----LIKPEMKIKLLMEGNVNGHQFVIEGDGKGHPFEGKQSMD   41
EosFP           1  MS----AIKPDMKINLRMEGNVNGHHFVIDGDGTGKPFEGKQSMD   41
KikG            1  MS----VITSEMKMELRMEGAVNGHKFVITGKGSGQPFEGIQNMD   41
KikGR           1  MS----VITSEMKIELRMEGAVNGHKFVITGKGSGQPFEGIQNVD   41
KikGRX          1  VS----VITSEMKIEVRMEGAVNGHKFVITGKGSGQPFEGIQNVD   41
dendGFP         1  MN----LIKEDMRVKVHMEGNVNGHAFVIEGEGKGKPYEGTQTAN   41
Dendra2         1  MN----LIKEDMRVKVHMEGNVNGHAFVIEGEGKGKPYEGTQTAN   41
Consensus_ss:              eeeeeeeeeee  eeeeeeeee      eeeeee

GFP            46  FICTTG-KLPVPWPTLVTTFSYGVQCFSRYPDHMKRHDFFKSAMP   89
ALLGFP_OLD_    42  LKVKEGAPLPFAYDILTTAFQYGNRVFTKYPKDIP--DYFKQSFP   84
ALLGFP_NEW_    42  LKVKEGAPLPFAYDILTTAFQYGNRVFTKYPKDIP--DYFKQSFP   84
ALLQ65H        42  LKVKEGAPLPFAYDILTTAFHYGNRVFTKYPKDIP--DYFKQSFP   84
LEAX6          42  LKVKEGAPLPFAYDILTTAFHYGNRVFAKYPKDIP--DYFKQSFP   84
LEA            42  LKVKEGAPLPFAYDILTTVFHYGNRVFAKYPKHIP--DYFKQSFP   84
LEAX72         42  LKVKEGAPLPFAYDILTTVFHYGNRVFTKYPKHIP--DYFKQSFP   84
LEAX121        42  LKVKEGAPLPFAYDILTTVFHYGNRVFAKYPKHIP--DYFKQSFP   84
Kaede          42  LVVKEGAPLPFAYDILTTAFHYGNRVFAKYPDHIP--DYFKQSFP   84
EosFP          42  LEVKEGGPLPFAFDILTTAFHYGNRVFAEYPDHIQ--DYFKQSFP   84
KikG           42  LTVIEGGPLPFAFDILTTVFDYGNRVFVKYPEEIV--DYFKQSFP   84
KikGR          42  LTVIEGGPLPFAFDILTTAFHYGNRVFVEYPEEIV--DYFKQSFP   84
KikGRX         42  LTVIEGGPLPFAFDILTTAFHYGNRVFVKYPEEIV--DYFKQSFP   84
dendGFP        42  LTVKEGAPLPFSYDILTTAVHYGNRVFTKYPEDIP--DYFKQSFP   84
Dendra2        42  LTVKEGAPLPFSYDILTTAVHYGNRVFTKYPEDIP--DYFKQSFP   84
Consensus_ss:      eeee        hhhhhhhh    eeeee        hhhhhh

GFP            90  EGYVQERTIFFKDDGNYKTRAEVKFEGD---TLVNRIELKGIDFK  131
ALLGFP_OLD_    85  EGYSWERSMTFEDGGICTATSDITLEGD---CFFYEIRFDGVNFP  126
ALLGFP_NEW_    85  EGYSWERSMTFEDGGICTATSDITLEGD---CFFYEIRFDGVNFP  126
ALLQ65H        85  EGYSWERSMTFEDGGICTATSDITLEGD---CFFYEIRFDGVNFP  126
LEAX6          85  EGYSWERSMTFEDGGICTARNDITLEGD---CFFNEIRFDGVNFP  126
LEA            85  EGYSWERSMTFEDGGICTARNDITLEGD---CFFNEIRFDGVNFP  126
LEAX72         85  EGYSWERSMTFEDGGICTARNDITLEGD---CFFNEIRFDGVNFP  126
LEAX121        85  EGYSWERSMTFEDGGICTARNDITLEGD---CFFYEIRFDGVNFP  126
Kaede          85  KGFSWERSLMFEDGGVCIATNDITLKGD---TFFNKVRFDGVNFP  126
EosFP          85  KGYSWERSLTFEDGGICIARNDITMEGD---TFYNKVRFHGVNFP  126
KikG           85  EGYSWERSMSYEDGGICLATNNITMKKDGSNCFVYEIRFDGVNFP  129
KikGR          85  EGYSWERSMSYEDGGICLATNNITMKKDGSNCFVNEIRFDGVNFP  129
KikGRX         85  EGYSWERSMSYEDGGICLATNNITMKKDGSNCFVNEIRFDGVNFP  129
dendGFP        85  EGYSWERTMTFEDKGICTIRSDISLEGD---CFFQNVRFNGMNFP  126
Dendra2        85  EGYSWERTMTFEDKGICTIRSDISLEGD---CFFQNVRFKGTNFP  126
Consensus_ss:        eeeeeeeeee   eeeeeeeeeee     eeeeeeeeeee
```

Figure 4. Sequence alignment of ALLGFP / LEA variants by PROMALS3D (*65*). The consensus predicted secondary structures are shown in the last line in each block. Letter ‘e’ represents β-sheet and ‘h’ does α-helix.

```
GFP           132  EDGNILGHKL-EYNYNSHNVYIMADKQKNGIKVNFKIRHNIEDGS  175
ALLGFP_OLD_   127  PNGPVMQKKTLKWEPSTEKMYVRD----GVLMGDVNMALLLEGGG  167
ALLGFP_NEW_   127  PNGPVMQKKTLKWEPSTEKMYVRD----GVLMGDVNMALLLEGGG  167
ALLQ65H       127  PNGPVMQKKTLKWEPSTEKMYVRD----GVLMGDVNMALLLEGGG  167
LEAX6         127  PNGPVMQKKTLKWEPSTEKMYVRD----GVLMGDVNMALLLEGGG  167
LEA           127  PNGPVMQKKTLKWEPSTEKMYVRD----GVLTGDINMALLLEGGG  167
LEAX72        127  PNGPVMQKKTLKWEPSTEKMYVRD----GVLTGDINMALLLEGGG  167
LEAX121       127  PNGPVMQKKTLKWEPSTEKMYVRD----GVLTGDINMALLLEGGG  167
Kaede         127  PNGPVMQKKTLKWEASTEKMYLRD----GVLTGDITMALLLKGDV  167
EosFP         127  ANGPVMQKKTLKWEPSTEKMYVRD----GVLTGDITMALLLEGNA  167
KikG          130  ANGPVMQRKTVKWEPSTEKMYVRD----GVLKGDVNMALLLQGGG  170
KikGR         130  ANGPVMQRKTVKWEPSTEKMYVRD----GVLKGDVNMALLLQGGG  170
KikGRX        130  ANGPVMQRKTVKWESSTEKMYVRD----GVLKGDVNMALLLQGGG  170
dendGFP       127  PNGPVMQKKTLKWEPSTEKLHVRD----GLLVGNINMALLLEGGG  167
Dendra2       127  PNGPVMQKKTLKWEPSTEKLHVRD----GLLVGNINMALLLEGGG  167
Consensus_ss:          hhh ee e    eeeeeee      eeeeeeeeeeee

GFP           176  VQLADHYQQNTPIGDG-PVLLPDNHYLSTQSALSKDPNEKRDHMV  219
ALLGFP_OLD_   168  HYRCDFK--TTYKAK-KGVQLPDYHFVDHRIEILSHDK-DYNNVK  208
ALLGFP_NEW_   168  HYRCDFK--TTYKAK-KGVQLPDYHFVDHRIEILSHDK-DYNNVK  208
ALLQ65H       168  HYRCDFK--TTYKAK-KGVQLPDYHFVDHRIEILSHDK-DYNNVK  208
LEAX6         168  HYRCDFK--TTYKAK-KGVQLPDYHFVDHCIEILSHDK-DYNNVK  208
LEA           168  HYRCDFK--TTYKAK-KGVQLPDYHFVDHCIEILSHDK-DYNNVK  208
LEAX72        168  HYRCDFK--TTYKAK-KGVQLPDYHFVDHCIEILSHDK-DYNNVK  208
LEAX121       168  HYRCDFK--TTYKAK-KGVQLPDYHFVDHCIEILSHDK-DYNNVK  208
Kaede         168  HYRCDFR--TTYKSRQEGVKLPGYHFVDHCISILRHDK-DYNEVK  209
EosFP         168  HYRCDFR--TTYKAKEKGVKLPGYHFVDHCIEILSHDK-DYNKVK  209
KikG          171  HYRCDFR--TTYKAK-KVVQLPDYHFVDHRIEITSHDK-DYNKVK  211
KikGR         171  HYRCDFR--TTYKAK-KVVQLPDYHFVDHQMEITSHDK-DYNKVK  211
KikGRX        171  HYRCDFR--TTYKAK-KVVQLPDYHFVDHLMEITSHDK-DYNKVK  211
dendGFP       168  HYLCDFK--TTYKAK-KVVQLPDYHFVDHRIEILGNDS-DYNKVK  208
Dendra2       168  HYLCDFK--TTYKAK-KVVQLPDAHFVDHRIEILGNDS-DYNKVK  208
Consensus_ss:      eeeeeeeeeeee            eeeeeeeeeeee      eee

GFP           220  LLEFVTA-AGITHGMDELYK-  238
ALLGFP_OLD_   209  LYEHAVA-RSS--LLPMTAA-  225
ALLGFP_NEW_   209  LYEHAVA-RYS--MLPRQAK-  225
ALLQ65H       209  LYEHAVA-RYS--MLPRQAK-  225
LEAX6         209  LYEHAVA--HS--GLPRQAK-  224
LEA           209  LYEHAVA-HSG---LPRQAK-  224
LEAX72        209  LYEHAVA-HSG---LPRQAK-  224
LEAX121       209  LYEHAVA-HSG---LPRQAK-  224
Kaede         210  LYEHAVA-HSG---LPDNVK-  225
EosFP         210  LYEHAVA-HSG---LPDNARR  226
KikG          212  LYEHAKA-HSG---LPRLAK-  227
KikGR         212  LYEHAKA-HSG---LPRLAK-  227
KikGRX        212  LYEHAKA-HSG---LPRLAK-  227
dendGFP       209  LYEHAVARYSP---LPSQAW-  225
Dendra2       209  LYEHAVARYSP---LPSQAW-  225
Consensus_ss:      eeeeeee ee
```

Figure 4. Continued

METHODS

Protein expression and purification

C-terminal 6 His-tagged reconstructed genes were inserted into the expression plasmid pGEM-T (Promega, Inc.). All plasmids were transformed into *E. coli* BL21(DE3) for expression. During the expression, proteins were induced with IPTG at OD_{600} of 0.6, the temperature was reduced from 37 ºC to 25 ºC, and cells were allowed to grow overnight. After the overnight culture, the cells were harvested with centrifugation at 8000 rpm for 15 min spin (Beckman) and pellets were stored at – 80 ºC. The protein purification started with sonication. The pellets were suspended in a lysis buffer, pH 7.9 50 mM N-cyclohexyl-2-aminoethanesulfonic acid (HEPES), 300 mM NaCl, 3 mM β-mercaptoethanol, 10% glycerol, 0.1 mM phenylmethylsulfonyl fluoride (PMSF), 1 mM $MgCl_2$, and DNase I. To avoid denaturation of the protein, the sonication was performed 5 cycles (1 minute run and 1 minute cooling, and amplitude 7; about 60% of power) on ice. The resulting cell lysate was clarified by centrifugation at 15000 rpm for 30 min at 4 ºC. The supernatant was collected and then centrifuged again at the same condition. After collecting the soluble faction, the proteins were purified by Ni-NTA (Qiagen) affinity chromatography. For a six-liter cell culture, 10 ml of Ni-NTA resin was used. Once the resin was settled in a column, it was washed with 10 times the bed volume of deionized water and 10 times of volume of column wash buffer (pH 7.9 50 mM HEPES, 300 mM NaCl). After the washing, the column was equilibrated with the supernatant and then washed again with the column wash buffer. Non-specifically bound proteins were eluted from the

column with a low imidazole buffer (pH 7.9 50 mM HEPES, 300 mM NaCl, and 20 mM Imidazole). The C-terminally His-tagged proteins were eluted with a high imidazole buffer (pH 7.9 50 mM HEPES, 300 mM NaCl, and 200 mM Imidazole) and collected with a fraction collector. The fractions from the high imidazole buffer were pooled and immediately dialyzed against a buffer containing pH 7.9, 50 mM HEPES, 20 mM NaCl and 1 mM ethylenediaminetetraacetic acid (EDTA) for overnight. After the dialysis, the protein solution was concentrated and flash-frozen. During expression, purification and storage, exposure to room light was minimized by wrapping all equipment, beakers, columns and trays in aluminum foil to prevent the photoconversion process. Floor shaker lid windows and glass refrigerator doors were obscured as well.

Site-directed mutagenesis for Q38 mutants

LEA Q38A, Q38N and Q38M mutations were introduced using QuickChange II Site-Directed Mutagenesis Kit (Agilent Technologies). All primers for Q38 mutants were designed using a web-based QuickChgange Primer Design Program available online (http://www.agilent.com/genomics/qcpd) (Table 1). To prepare the sample, 5 μl of 10X reaction buffer, 1 μl of LEA plasmid (50 ng/μl), 1 μl of forward primer, 1 μl of reverse primer, 1 μl of dNTP mix, 3 μl of Quiksolution, and 40 μl of ddH2O then 1 μl of *PfuUltra* HF DNA polymerase (2.5 U/μl) was added in a PCR tube. The thermal cycling reaction was performed using TechGene Thermal Cycler FTgene2D (Techne, Burlington, NJ) with the following condition; 1. 95 ^{o}C for 5 minutes, 2. 95 ^{o}C for 50 seconds, 60 ^{o}C for 50 seconds, and 68 ^{o}C for 5 minutes (repeat 18 times), and 3. 68 ^{o}C for 7 minutes.

Table 1

Primer sequences for Gln42(38) mutations for LEA

	Primer Sequence	Length (nt.)	T_m (^oC)
Q38A	5’ – caaaccgtatgaaggcaccgcgaccatgaacctgaaagtg – 3’	40	80.13
	5’ – cactttcaggttcatggtcgcggtgccttcatacggtttg – 3’	40	80.13
Q38N	5’ – caaaccgtatgaaggcaccaataccatgaacctgaaagtga – 3’	41	78.16
	5’ – tcactttcaggttcatggtattggtgccttcatacggtttg – 3’	41	78.16
Q38 M	5’ – caaaccgtatgaaggcaccatgaccatgaacctgaaagtg – 3’	40	79.10
	5’ – cactttcaggttcatggtcatggtgccttcatacggtttg – 3’	40	79.10

Following the cycling reaction, the sample tubes were put on ice for 2 minutes to cool the reaction to 37 ^{o}C. After the thermal cycling reaction, 1 μl of *Dpn* I restriction enzyme (10 U/μl) was added directly to each sample. The samples were gently mixed and spun down in a microcentrifuge for 1 min and then incubated at 37 ^{o}C for 1 hour to digest nonmutated plasmid. After the *Dpn* I digestion, the mutants were transformed into *E. coli* XL10-Gold and grown overnight on LB / carbenicillin plates. For DNA sequencing, colonies of interest were cultured overnight in 5 ml of LB/carbenicillin plate and then the plasmids were purified using QIAprep Miniprep (QIAGEN).

Crystallization and data collection.

Primary screens for crystallization were carried out with Hampton Crystal Screen Cryo (Hampton Research) and previous other GFP-like proteins crystallization conditions. Crystals were grown by hanging drop vapor diffusion at 4 ^{o}C (ALL-GFP) or 20 ^{o}C (all others) in the dark. The drops consisted of 2 μl of protein (10 mg/ml in 50 mM HEPES pH 7.9, 20 mM NaCl and 1 mM EDTA) and 2 μL of mother liquor. For ALL-GFP crystal growth, the mother liquor consisted of 60 mM Tris (pH 7.7), 100 mM magnesium sulfate 20 % PEG 4000; for ALL-Q65H, 200 mM Tris (pH 7.7), 120 mM magnesium sulfate, 20 % PEG 4000; for LEA-X6, 70 mM Tris (pH 7.7), 200 mM magnesium chloride, 16 % PEG 4000; for LEA-X121, 70 mM Tris-HCl (pH 7.5), 100 mM magnesium sulfate, 14% PEG 4000, 15% glycerol; and for LEA, 70 mM Tris (pH 7.6), 100 mM Sodium Acetate, 12 % PEG 4000, 15 % glycerol; for LEA Q38A 85 mM Tris (pH 7.5), 85 mM Sodium Acetate, 16 % PEG 4000.

Crystals were flash-frozen in liquid nitrogen, and all X-ray diffraction data were collected at 100 °K. Most data sets were collected using an R-AXIS IV++ image plate detector mounted on a Rigaku RU200HB rotating anode generator equipped with osmic confocal mirrors. Data sets of LEA and LEA-X121 were collected at the Advanced Photon Source (Argonne National Laboratory) beamline ID-19, using an ADSC Q315R Berkeley National Laboratory) beamline 4.2.2 using a NOIR-1 detector (Table 2).

Structure determination and refinement.

ALL-GFP: Indexing and integration were carried out using MOSFLM (*66*) and the data were scaled and merged with SCALA in the CCP4 suite of programs (*67*) The structure of ALL-GFP was solved by molecular replacement using the program PHASER (*68*), and the A-chain of KikG, from the coral *Favia favus*, as a search model (PDB ID 1XSS) (*38*). All four chains in the asymmetric unit were located, and the initial electron density was auto-traced using ARP/wARP (*69*) to introduce the correct ALL-GFP amino acid sequence. Side chain positions were adjusted manually within the program COOT (*70*). After 5% of the data were removed for R_{free} calculations, tetramer rigid body refinement was carried out using the program REFMAC (*71*). Positional refinement was initially carried out in the absence of a chromophore model, while restraining the four chains of the asymmetric unit to maintain non-crystallographic symmetry. After several rounds of refinement to the full resolution limit, the chromophore was modeled by using the chromophore coordinates from the cmFP512 X-ray structure (PDB ID 2C9J) (*72*). I chose this model based on the high resolution (1.3 Å) of cmFP512, which

bears a chromophore derived from the Gln-Tyr-Gly (QYG) sequence as does ALL-GFP. The chromophore was subsequently refined against an appropriate target geometry file as described elsewhere (*46*). At this point, all non-crystallographic symmetry (NCS) restraints were eliminated, and isotropic B-factor refinement was carried out prior to water modeling. Subsequently, solvent molecules were modeled into strong electron density ($2F_O$-F_C one sigma or higher) within hydrogen bonding distance to protein groups. Final rounds of refinement were carried out in REFMAC with inclusion of solvent molecules, all NCS restraints lifted, and with anisotropic B-factor refinement.

ALL-Q65H, LEA-X6, LEA X-121: These structures were solved by molecular replacement using the A-chain of ALL-GFP (Old) as a search model. The chromophores were modeled by using coordinates from ALL-GFP (Old), edited to reflect the Q65H substitution to model the chromophore of ALL-Q65H, the coordinates of Kaede (PDB ID 2GW3) (*51*) to model the LEA chromophore, and the coordinates of LEA to model the LEA-X6 and LEA-X121 chromophores. All other procedures were carried out as described above. As the ALL-Q65H structure bears a tetramer per asymmetric unit, the model was initially refined with NCS restraint. However, these were lifted during later stages of refinement.

LEA: The LEA data set was processed using the HKL2000 software package (*73*). Molecular replacement was carried out with the program PHASER (*68*), using the X-ray structure of the A-chain of EosFP (PDB ID code 1ZUX) (*50*) as a search model. The initial electron density was auto-traced using the BUCCANEER software to introduce the correct LEA amino acid sequence. The

Table 2

Crystallographic data collection statistics

	ALL-GFP (OLD)	ALL Q65H	LEAX(6)	LEA	LEA X121	LEA Q38A
space group	C2	C2	I222	I222	I222	I222
unit cell						
a/b/c (Å)	106.49/93.06/ 88.99	106.28/92.71/ 89.06	45.65/78.34/ 119.43	48.77/79.02/ 117.89	49.04/79.18/ 118.04	49.99/81.14/ 117.80
$\alpha/\beta/\gamma$ (°)	90/92.72/90	90/93.18/90	90/90/90	90/90/90	90/90/90	90/90/90
Detector	R-AXIS IV++	NOIR-1	R-AXIS IV ++	ADSC Q315	ADSC Q315	R-AXIS IV ++
Wavelength (Å)	1.5418	0.9998	1.5418	0.9792	0.9792	1.5418
Resolution (Å)	28.35 – 1.60	29.67 – 1.40	39.16 – 2.50	50.00 – 1.85	34.05 – 1.75	34.50 – 1.95
High-resolution shell (Å)	1.69 – 1.60	1.48 – 1.40	2.64 – 2.50	1.88 – 1.85	1.84 – 1.75	2.06 – 1.95
Total observations	264,463	598,321	39,140	73,879	84,219	101,622
Unique reflections	111,934	169,035	7,913	19,903	23,358	17,887
Redundancy	2.4 (2.1)	3.5 (3.5)	4.8 (4.9)	3.7 (3.7)	3.6 (3.7)	5.7 (5.7)
Av I/σ	13.4 (1.8)	6.1 (1.8)	16.2 (4.1)	6.8 (4.1)	7.6 (2.3)	8.6 (2.2)
Completeness (%)	98.2 (92.6)	99.9 (100.0)	98.8 (98.4)	99.6 (99.6)	99.3 (99.9)	100.0 (100.0)
$R_{merge\ (\%)}$ [b]	4.7(42.8)	7.6 (37.8)	5.8 (36.3)	7.2 (37.3)	8.2 (38.1)	6.0 (35.0)

[a] Values within parentheses refer to the high resolution shell

[b] $R_{merge} = \Sigma|I_{hkl} - \langle I \rangle| / \Sigma\langle I \rangle$, where $\langle I \rangle$ is the average of individual measurements of I_{hkl}.

Table 3

Crystallographic refinement statistics

	ALL-GFP (OLD)	ALL Q65H	LEAX(6)	LEA	LEA X121	LEA Q38A
Resolution range (Å)	28.12 – 1.60	29.67 – 1.40	30.55 – 2.50	41. 50 – 1.85	34.05 – 1.75	30.88– 1.95
Number of reflections	106,316	160,594	7,527	18,802	22,027	17,266
R_{cryst} (%)[c]	17.3	14.3	19.7	15.1	17.1	18.49
R_{free} (%)[c]	21.0	17.9	24.7	19.5	22.6	23.77
rmsd bond lengths (Å)	0.011	0.011	0.018	0.014	0.016	0.022
rmsd bond angles(°)	1.402	1.492	1.758	1.519	1.629	1.818
average B factors						
all atoms (Å^2)	11.434	10.361	31.523	22.869	20.813	29.461
Total atoms	8,082	8,576	1,775	1,905	1,960	1,806

[c] R_{cryst} and $R_{free} = \Sigma_h (||F(h)_{obs}| - |F(h)_{calc}||) / |F(h)_{obs}|$ for reflections in the working and test sets (5% of all data).

chromophore was modeled by use of the LEA chromophore coordinates. All other computational procedures were carried out as described for ALL-GFP. The data processing and refinement statistics are shown in Table 2 and 3.

LEA Q38A: The LEA Q38A data set was processed using MOSFLM (*66*) and scaled and merged with SCALA in the CCP4 suite of programs (*67*). Molecular replacement was carried out with the program PHASER (*68*), using the X-ray structure of the LEA protein as a search model. After the initial refinement, Gln-38 was replaced with Ala using COOT (*70*). The chromophore was modeled by use of the LEA chromophore coordinates. All other computational procedures were carried out as described for ALL-GFP.

Absorbance, fluorescence, and chromophore pKa determinations.

Absorbance spectra were collected on a Shimadzu UV-2401 spectrophotometer from pH 4.0 to 13.0 at room temperature. Protein stock solutions (1 mg/ml) were diluted 10-fold into appropriate buffers and allowed to equilibrate at a concentration of 0.1 mg/mL. pH 4.0 to 11.0 buffers consisted of 50 mM concentration of citrate, acetate, piperazine-1, 4-bis(2-ethanesulfonic acid) (PIPES), HEPES, *N*-cyclohexyl-2-aminoethanesulfonic acid (CHES), or *N*-cyclohexyl-3-aminopropanesulfonic acid (CAPS), all containing 100 mM NaCl and 1 mM EDTA. Buffers at pH 12 and 13 were prepared by adjusting NaOH concentrations to 10 and 100 mM. The chromophore pKa of each variant was determined by curve fitting the absorbance intensity of the chromophore anion to

the Henderson-Hasselbalch equation using the software package Kaleidagraph (Synergy). The data were fit to the expression

$$A = A_{max}/[1 + 10^{(pKa-pH)nH}] \quad (1)$$

, where A_{max} represents the absorbance upon full ionization, K_a represents the apparent proton dissociation constant for the titrating site(s), and n represents the Hill coefficient n_H.

Fluorescence excitation and emission spectra were collected using a SPEX Fluoromax-3 (Jobin Yvon Horiba) fluorometer with integration time set to 0.1 S, and excitation and emission slit width set to 1.0 nm. The 1 mg/ml protein samples were kept at pH 7.9 in 50 mM HEPES, 20 mM NaCl and 1 mM EDTA buffer. For measurements, the excitation and emission wavelengths were in table X.

Quantum yield determinations.

The quantum yields of green fluorescence Φ_{Fl} of All-GFP/LEA variants were determined using fluorescein as a standard (1 μM in 0.1 M NaOH). The buffer for all protein preparations was 50 mM Hepes pH 7.0, 50 mM NaCl, and 1 mM EDTA. A new fluorescein standard was prepared for each protein. Briefly, the cross-over point of protein and fluorescein absorbance spectra was utilized as the excitation wavelength for fluorescence emission scans collected on the same samples (integration time 1.0 s; increment 1.0 nm; slit width 1.0 nm). The corrected emission scans were integrated with the left integration limit set to the

wavelength at which 1/3 of maximum fluorescence intensity was observed. The integration results were fit to the following equation: $\Phi_{Fl} = \Phi_R(I/I_R)(OD_R/OD)$, where Φ_{Fl} is the quantum yield, I is the integrated intensity, and OD is the optical density at the respective excitation wavelength. The subscript "R" stands for the reference fluorophore (fluorescein) with known quantum yield of 0.95.

Photoconversion efficiency determination

Photoconversion reaction samples (0.1 mg/mL) were prepared with 100 μL of 1 mg/mL of protein stock solutions and 900 μL of each buffers and incubated overnight in the dark. After UV exposure, absorbance of each sample was measured from 250 to 650 nm using Shimadzu UV-2401 spectrophotometer. The baseline was corrected with the appropriate buffer as a blank before the measurement. The progression of the photoconversion reaction was measured as absorbance change of total green chromophore. The total absorbance of the green chromophore was calculated using the values at 386 nm (neutral chromophore) and 504 nm (anion chromophore). For the anionic chromophore, the molar extinction coefficient is 86,754 M^{-1} cm^{-1}. Based on the molar coefficient of the anion chromophore, an extinction coefficient for the neutral chromophore was calculated using the absorbance intensity at pH 6.34, which is the pKa value of green LEA chromophore. At this pH value, the number of neutral and anionic chromophores is the same. Therefore, I can use this equation;

$$Abs_{570nm}/\varepsilon_{anion} = Abs_{370nm}/\varepsilon_{neutral} \qquad (2)$$

, where ε represents the molar extinction coefficient and Abs represents absorbance intensity. From the calculation, the molar extinction coefficient of neutral chromophore is 22943 M^{-1} cm^{-1}. With both molar extinction coefficients, the absorbance values of the neutral and anionic chromophores were converted into the concentration of total green chromophore.

The concentration of total chromophore was plotted as a function of time and fit to a first-order reaction (equation 3), using the software Kaleidagraph (Synergy). Below pH 7.0, a large decrease was observed between 0 and 1 minute. However the data starting with 1 minute fit well to a first-order reaction,

$$A_t = A_0 e^{-kt} \quad (3)$$

, where A_0 represents the absorbance of total green chromophore at 0 minute, and k represents the observed rate constant (k_{obs}). The resulting k_{obs} values were plotted as a function of pH, and the apparent Ka values were determined by fitting to equation 4 with the software Kaleidagraph (Synergy).

$$k_{obs} = \frac{k_{int} K_{a1} [H^+]}{K_{a1} [H^+] + [H^+]^2 + K_{a1}K_{a2}} \quad (4)$$

This equation is derived from a simple diprotic system where the two successive acid dissociation constants are closer than 3.5 pH units, which described in 'Enzyme Kinetics' by Segel (pp. 914-917) (*74*).

Table 4

Buffers for the photoconversion reactions

pH	Buffer
4.03, 4.50, 4.76, 5.00, 5.14, 5.40, & 5.58	100 mM Sodium Acetate, 200 mM NaCl, and 2 mM EDTA
6.15 & 6.33	100 mM PIPES, 200 mM NaCl, and 2 mM EDTA
6.45, 6.63, 6.84, & 7.01	100 mM Bis-Tris, 200 mM NaCl, and 2 mM EDTA

Results

To study the structural requirements for green-to-red photoconversion of Kaede-type proteins, several high-resolution X-ray structures of reconstructed ancestral proteins were solved. The amino acid sequences of these proteins are based on a statistical analysis of extant GFP-like proteins that belong to the suborder *Faviina* along clade D of the stony corals. The reconstructed protein ALL-GFP is thought to present a deep ancestral node, the common green ancestor from which color diversification has evolved along the *Faviina* suborder (Figure 2) (*20*). I have termed this protein the ALL-GFP (old) ancestor, as the sequence of the C-terminal clasp has since been redesigned based on a more recent statistical analysis that included newly available sequences. The redesigned ALL protein has been termed ALL-GFP (new) and is considered an improved representation of the common green ancestor. The modifications concern residues 217 to 225 (ALL residue numbering system, see Figure 4 sequence alignment), the protein's C-terminus that wraps around its neighboring protomer to form a tight tetrameric assembly (Figure 8), as originally described for DsRed (*75*). The redesigned C-terminus was incorporated into all subsequent clones (Figure 4). The reconstructed Kaede-like FP is called LEA (Least Evolved Ancestor). LEA protein has 13 mutations from the ALL-GFP ancestor and shows the green-to-red photoconversion reaction with UV light. In this dissertation, I will use the consensus GFP numbering system and provide the ALL protein residue numbers in parenthesis (Table 5). As all proteins were expressed and purified in the dark, and all crystal growth was carried out in the dark, the crystal structures presented here are of proteins that have matured to the green-fluorescent state, but have

Table 5

LEA mutations

GFP Residue number	ALL-GFP/LEA Residue number	ALL-GFP	LEA X(6)	LEA
63	60	Ala	Ala	Val
65	62	Gln	His	His
72	69	Thr	Ala	Ala
77	74	Asp	Asp	His
109	104	Thr	Arg	Arg
110	105	Ser	Asn	Asn
121	116	Tyr	Asn	Asn
162	154	Met	Met	Thr
165	157	Val	Val	Ile
204	194	Arg	Cys	Cys
227	216	Arg	His	His
-	217	Tyr	Deletion	Deletion
228	219	Met	Gly	Gly

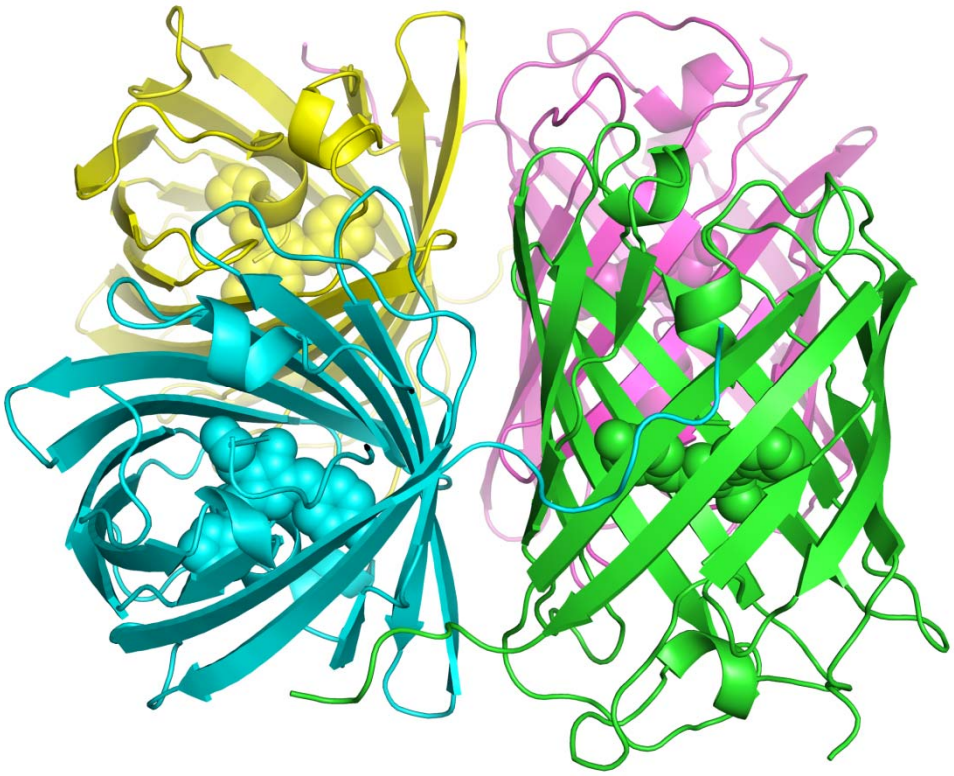
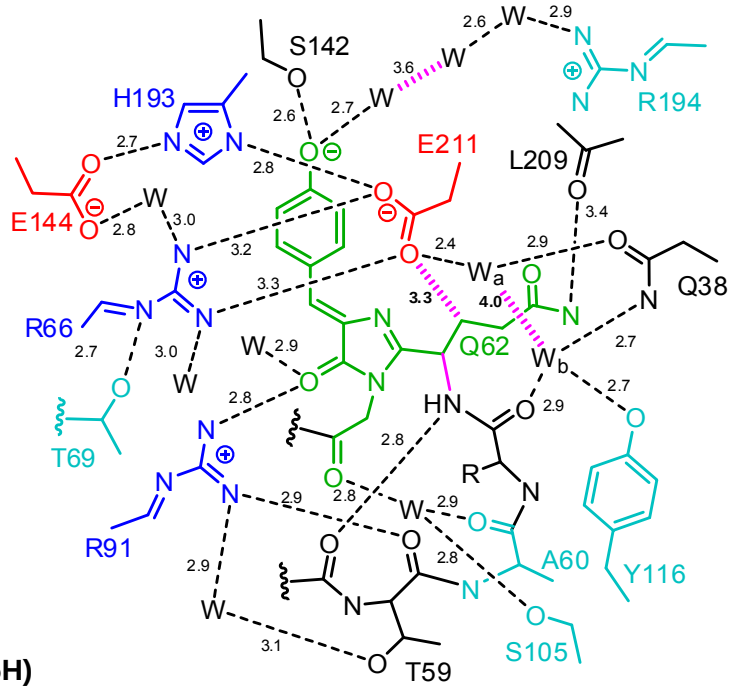


Figure 5.Tetrameric assembly of ALL Q65H. Overall structure is shown in ribbon representation and the chromophore is shown in sphere.

A (ALL-GFP)



B (ALL Q65H)

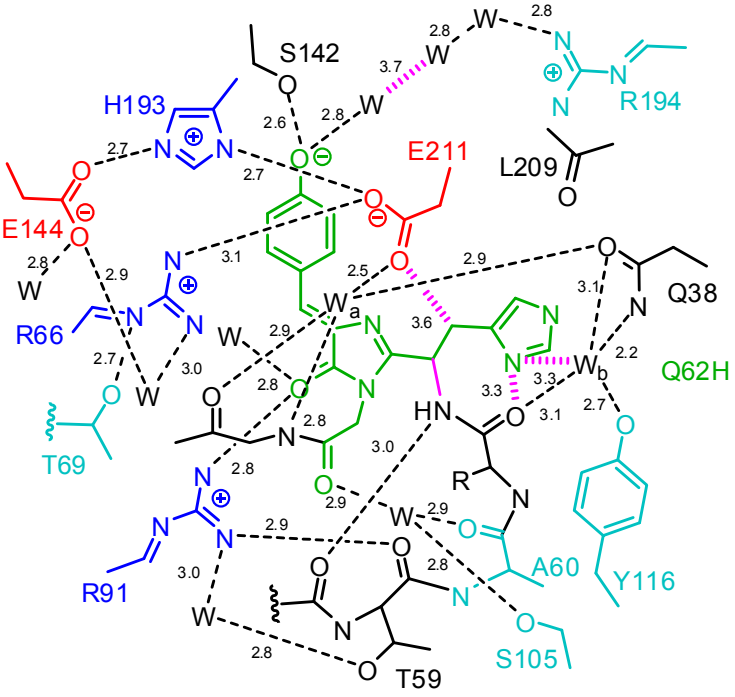
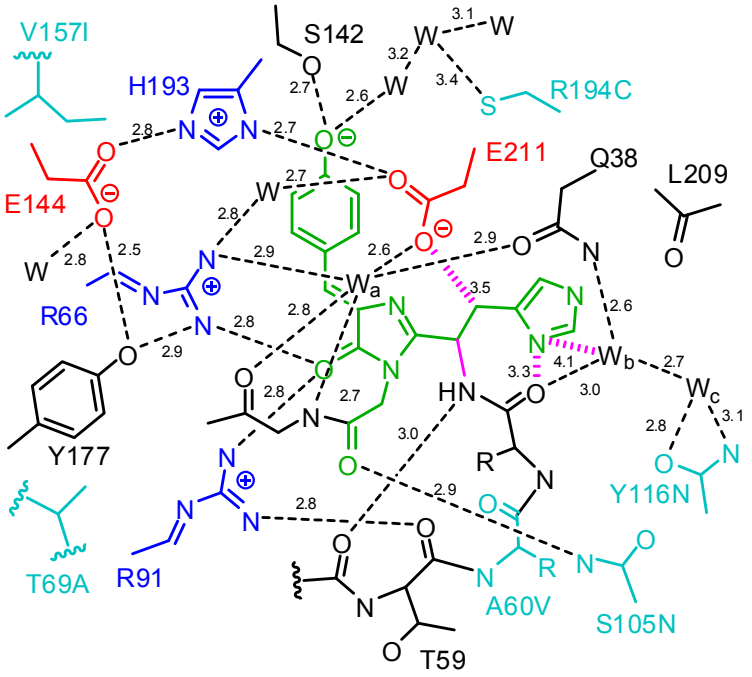


Figure 6. Schematic representation of ALL-GFP / LEA variants chromophore and environment - Inferred hydrogen bonds are shown as dashed lines. The chromophore is shown in green. Non-hydrogen-bond distances as pink bars. (following ALL-GFP residue numbers)

C (LEA)



D (LEA X6)

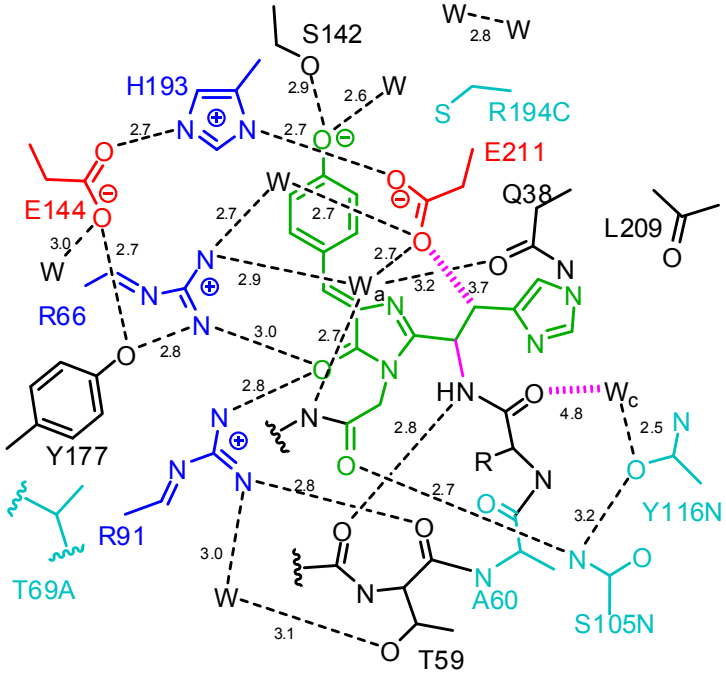


Figure 6. continued

E (LEA X121)

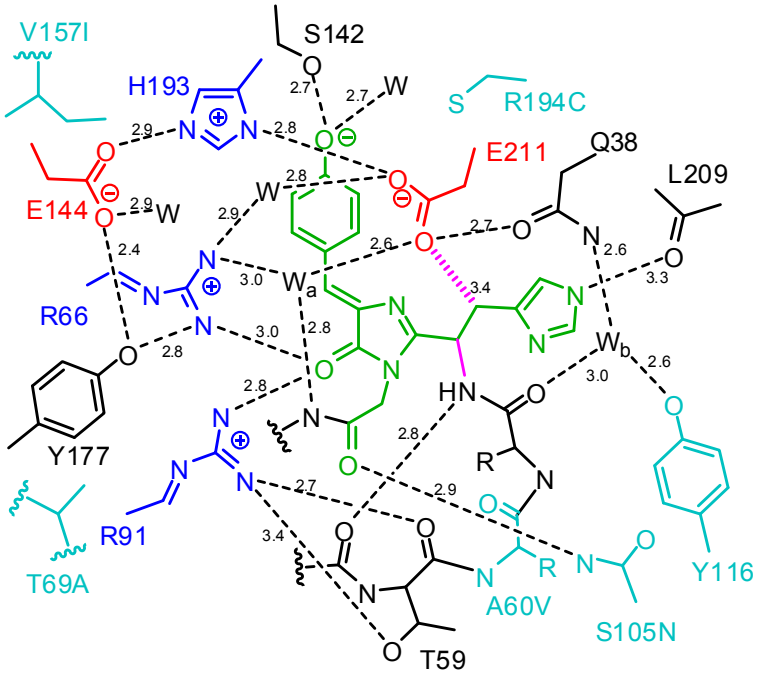


Figure 6. continued

not undergone any photoconversion to the red state. Therefore, at the crystallization pH of 7.5 to 7.7, which was at least 1.3 pH units above the pKa of the chromophore, the observed absorbance bands are centered on 489 to 506 nm, with emission maxima between 502 and 514 nm.

Crystal structures of the ALL-GFP / LEA variants

I have obtained a 1.6 Å resolution X-ray structure of the ALL-GFP ancestor, a reconstructed protein originally described by Ugalde *et al*. (*20*). The ALL-GFP (old) protein has four chains per asymmetric unit, providing four independent views of the chromophore and its environment. Superimposition of the B, C, and D chains onto the A-chain using a least squares fit algorithm as implemented in the program COOT (*70*) provided root-mean-square deviations (RMSD) of 0.17 to 0.20 Å for the α-carbons, and an RMSD of 0.48 to 0.57 Å for all atoms. Likewise, the secondary structure matching algorithm SSM (*76*) provided an RMSD of 0.17 to 0.20 Å (Table 6). Importantly, the geometry of the chromophore and surrounding residues was found to be nearly identical in all four chains. In addition, I have solved the structure of ALL-Q65H to 1.4 Å resolution. The Q65H substitution, absolutely required for red color acquisition, was introduced into the ALL-GFP (new) background. This protein also crystallized in spacegroup C2 with four molecules per asymmetric unit, and pair-wise least-squares superimpositions of chains A, B, C, and D provided an RMSD ranging from 0.17 to 0.19 Å for α-carbons, and 0.49 to 0.67 Å for all atoms (Table 6). I have solved the X-ray structure of LEA to 1.8 Å resolution and those of the LEA variants LEA-X6 to 2.5 Å, LEA-X121 to 1.7 Å, and LEA Q38A 1.9 Å. Crystals of LEA-X72 were

also obtained, although the data sets were of inferior quality. LEA crystallized in space group I222 with one molecule per asymmetric unit, and bears the set of mutations both necessary and sufficient for the efficient acquisition of red color. Compared to ALL-GFP (new), LEA bears the 12 substitutions and one deletion; A63V, Q65H, T72A, T109R, S110N, Y121N, M162T, V165I, R204C, R227H, delY[227-228], and M228G (GFP consensus numbering) (*62*). Five sites concern residues with side chains clustering near the chromophore, A63V, Q65H, S110N, Y121N, and V165I (Figure 7, red residues) (*62*). The remaining 8 sites concern residues with solvent-exposed side chains, or residues that are part of the C-terminal clasp (Figure 7, blue residues). LEA-X6, LEA-X121, and LEA-X72 are thought to be transitional variants, possible evolutionary intermediates on the pathway of phenotypic change from green to Kaede-type red. However, the statistical analysis does not allow for a prediction of the sequence of events, i.e. which mutation was introduced first along a particular evolutionary branch, or whether parallel evolution played a role. The specific residues of LEA that were mutated to the original residues present in ALL-GFP were chosen for further investigation because they showed a large effect on the red color after the reverse mutagenesis (*62*). LEA-X6 is a transitional variant where four LEA residues were reverted to the respective ALL-GFP residue: V63(60)A, H77(74)D, T162(154)M, and I165(157)V and these are among the 'fine tuning' residues, which display lesser impact to red fluorescent (*62*). LEA-X121 bears the N121(116)Y reversion, and LEA-X72 the A72(69)T reversion. Pair-wise superimpositions of LEA variants and ALL-GFP(old) using a secondary structure matching algorithm (*76*) provided an α-carbon root-mean-square deviations (RMSD) ranging from 0.14 to 0.29 Å, indicating only minimal structural rearrangements (Table 6).

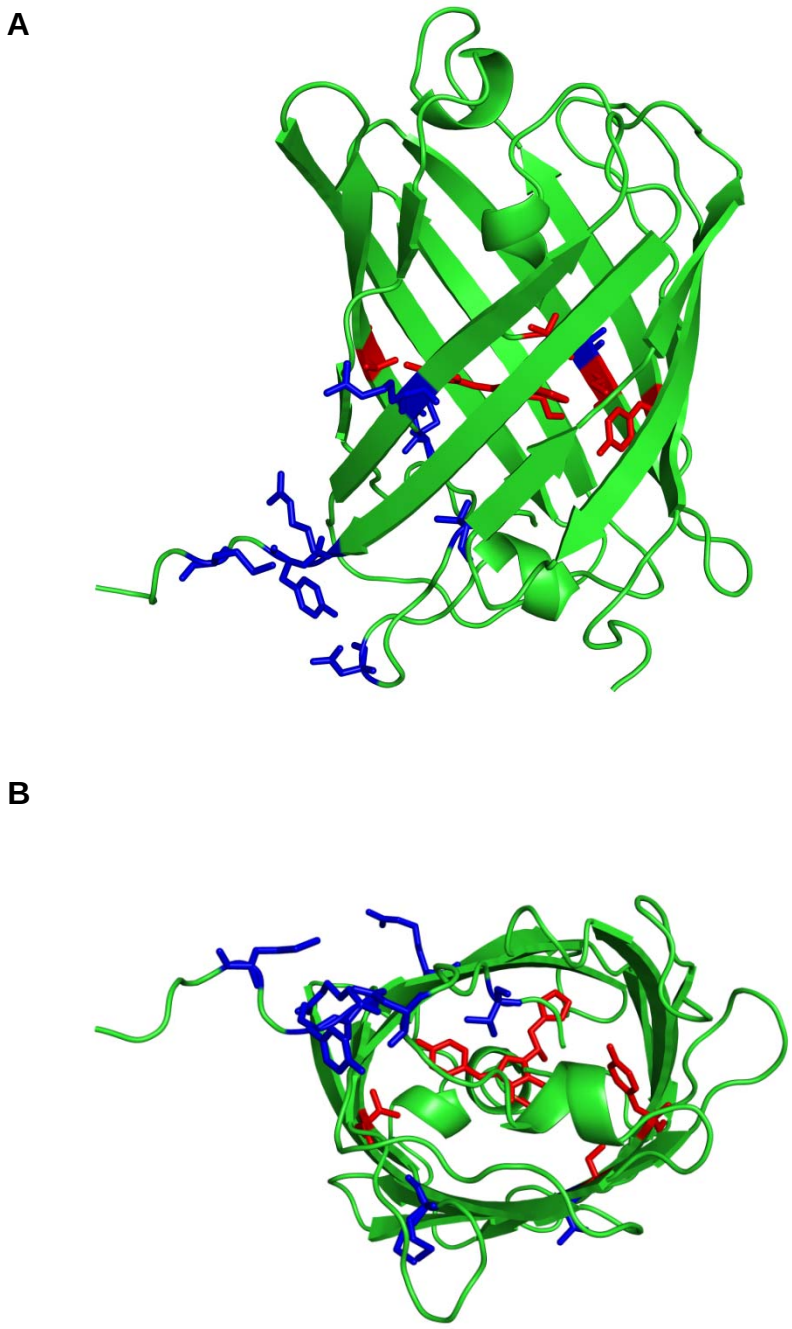


Figure 7. LEA mutations overlay on ALL Q65H crystal structure - Red residues (A63V, Q65H, S110N, Y121N, and V165I) represents the mutations around the chromophore and blue residues (T72A, D77H, T109R, M162T, R204C, R227H, delY227, M228G) are residues with solvent-exposed or part of the C-terminal chain.

Table 6.

C_α RMSD values from (A) SSM (Secondary Structure Matching) Superpose and (B) LSQ (Least Square Fit) superpose.

A

	ALL-GFP	ALL Q65H	LEA X6	LEA	LEA X121	LEA Q38A	Kaede	EosFP	Dendra2
ALL-GFP		0.20	0.29	0.30	0.36	0.40	0.58	0.53	0.33
ALL Q65H	0.20		0.28	0.26	0.21	0.40	0.59	0.54	0.36
LEA X6	0.29	0.28		0.27	0.26	0.23	0.50	0.46	0.38
LEA	0.30	0.26	0.27		0.14	0.20	0.47	0.44	0.38
LEA X121	0.36	0.21	0.26	0.14		0.21	0.51	0.46	0.43
LEA Q38A	0.40	0.40	0.23	0.20	0.21		0.53	0.47	0.47
Kaede	0.58	0.59	0.50	0.47	0.51	0.53		0.47	0.53
EosFP	0.53	0.54	0.46	0.44	0.46	0.47	0.47		0.58
Dendra2	0.33	0.36	0.38	0.38	0.43	0.47	0.53	0.58	

B

	ALL-GFP	ALL Q65H	LEA X6	LEA	LEA X121	LEA Q38A	Kaede	EosFP	Dendra2
ALL-GFP		0.20	0.30	0.30	0.37	0.40	1.70	1.68	0.33
ALL Q65H	0.20		0.28	0.26	0.36	0.41	1.73	1.82	0.36
LEA X6	0.29	0.28		0.26	0.26	0.23	1.66	1.66	0.37
LEA	0.30	0.27	0.26		0.14	0.20	1.63	1.64	0.38
LEA X121	0.37	0.36	0.26	0.14		0.21	1.68	1.67	0.43
LEA Q38A	0.40	0.41	0.23	0.20	0.21		1.69	1.67	0.46
Kaede	1.70	1.73	1.66	1.63	1.68	1.69		070	1.66
EosFP	1.68	1.82	1.66	1.64	1.67	1.67	0.70		1.77
Dendra2	0.33	0.36	0.37	0.38	0.46	0.46	1.66	1.77	

Each protomer has a typical GFP-like protein structure, which is an 11-stranded β-barrel with a central α-helix and the chromophore is found in the center of the barrel. Compared to Kaede-type proteins, ALL-GFP / LEA variants have quite similar structure with RMSD between all C_α atoms of 1.7 Å (Table 6). However, ALL-GFP and ALL Q65H have higher average B-factor of the model than LEA variants. LEA X6 and LEA Q38A showed the highest B-factor values among of them. The orientation of the tetrameric structures is almost identical to those of EosFP and Kaede. Like other Anthozoan FPs, it is held together by two different interfaces – A/B perpendicular & A/D antiparallel orientation. The A/B interface is dominated by hydrophilic interaction. In contrast to A/B interface, the interaction between A and D is mainly hydrophobic interactions. In the A/B interface, a few salt bridges exist including Glu96/Arg149, Asp156/Arg170, and Asn158/Lys145. Additionally, there is a hydrogen-bonded cluster consisting of Thr143, Arg170, and Asp156. Also Tyr188, Phe190, and His212 provide a hydrophobic patch (Figure 8). The electron density omit maps of the green chromophores, 4-(*p*-hydroxybenzylidene)-5-imidazolinone, of all structures solved are shown in Figure 9. The chromophores demonstrate that a high degree of co-planarity exists between the 2-ring GFP-like chromophore consisting of the imidazolinone and the phenolic ring conjugated by a β-methylene bridge. Torsion angles of the chromophores are around 160-170 degrees (Table 7). As observed previously (*50, 51*), no major structural rearrangements are necessary to generate the red chromophore, where the His65 imidazole ring is found to be π-conjugated to the GFP-like chromophore (Figure 3). Also there is no distinct hydrogen bond network from the phenolic ring to the imidazole of His65(62).

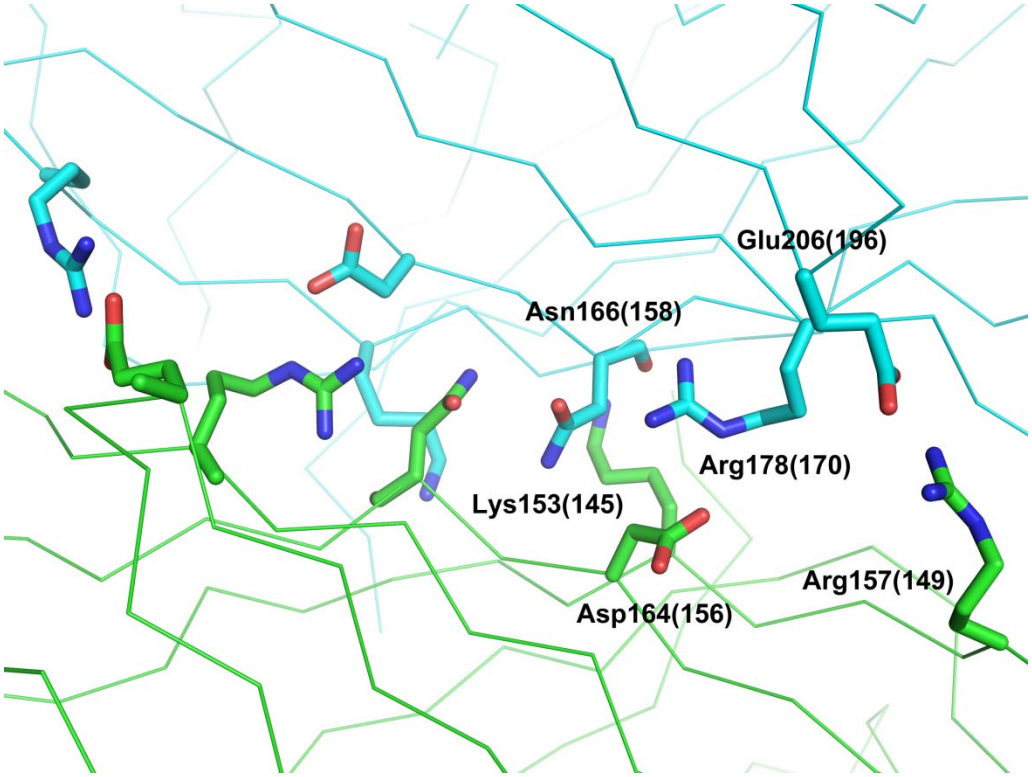


Figure 8. The dimer interface between A and B (C and D) chains. Each chain is shown as C_α traces in green or cyan and the amino acids responsible for the interactions are shown as stick presentation.

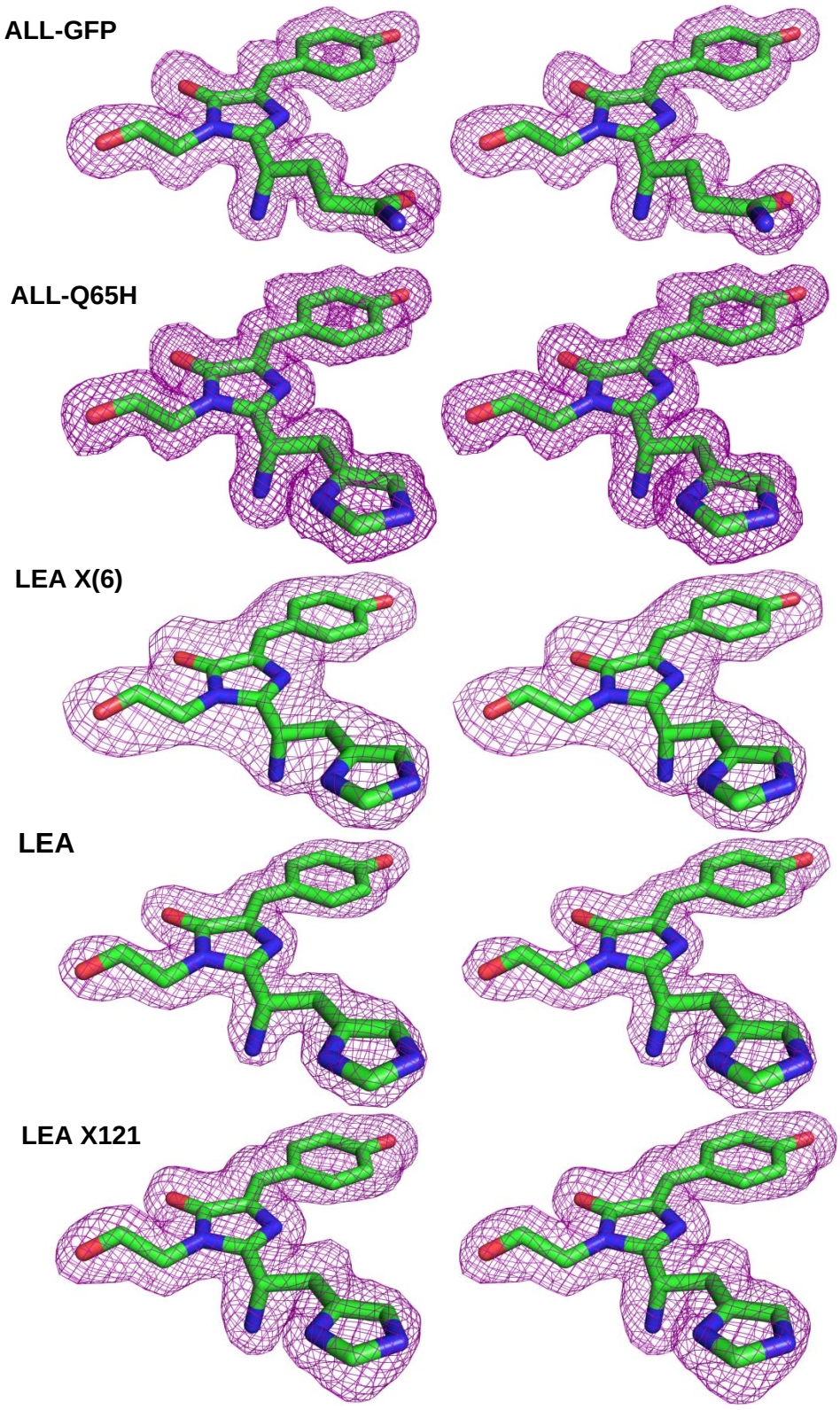


Figure 9. Stereo view of the $F_o - F_c$ electron density maps of the green chromophores

Table 7

Torsion angles between the GFP-like chromophore's phenol ring and the imidazolinone ring

Proteins	Torsion angles (degree)
ALL-GFP	170.71
ALL Q65H	168.72
LEA X(6)	159.93
LEA	164.75
LEA X121	162.96
LEA Q38A	159.38

H-bonded network for proton transfer from Glu211 to carbonyl oxygen of the leaving group

The carboxylate of Glu211 may serve as a general base in initiating the chemical steps upon photoexcitation. I have carried out extensive comparisons of the arrangement of functional groups and hydrogen-bonded networks in the vicinity of the green chromophores of the six variants under investigation. As expected, we find that the carboxylate of Glu222(211) is always positioned within van der Waals distance of the β-carbon of His65(62) (Table 8). Glu222(211) is believed to act as general base in a proton abstraction step from this carbon, thus facilitating full conjugation of the imidazole group with the GFP-like chromophore. However, in the photoconvertible set of proteins, two additional ordered water molecules are located between Arg69(66) and Glu222(211), whereas in the non-photoconvertible ALL-GFP and ALL Q65H, these waters are absent. Instead, the guanidinium group of Arg69(66) and the carboxylate of Glu222(211) are directly hydrogen bonded to each other via two H-bonds (Figure 10). The additional waters may bear significance in terms of reducing the energetic cost of Glu222(211) protonation. In the non-photoconvertible ALL Q65H, neutralizing the carboxylate of Glu222(211) would require breakage of the salt bridge to Arg69(66), an energetically costly event. In addition, I find that in all structures of sufficient resolution (which excludes LEA X(6) at only 2.5 Å), a hydrogen bonded network leads from the Glu222(211) carboxylate to the carbonyl oxygen of the leaving group that partakes in bond scission. This network includes a water and Gln42(38). Although a glutamine is an unusual amino acid residue to participate in proton transfer reactions, it cannot be

Table 8

The distances of key position in ALL-GFP and LEA variants

Protein	Glu211 Carboxylate Group – C_β His62(Å)	His62-N_δ - Phe61 carbonyl Group (Å)
ALL-GFP	3.4	-
ALL Q65H	3.6	3.3
LEAX(6)	3.5	3.4
LEA	3.7	3.3
LEAX121	3.4	3.4
LEA Q38A	3.8	-

completely excluded. Therefore, this network is thought to provide structural stability without active participation in the chemical steps of photoconversion. In the non-convertible Q65H variant, one water is H-bonded to the carboxamide leaving group, whereas in LEA (highly photoconvertible), two waters are positioned near the leaving group. This feature may be significant as they may serve as proton donors to the carboxamide leaving group.

Reorganization of interior packing and optimization of hydrogen bonding.

The substitutions Y116N, A60V, and S105N serve to subtly improve packing arrangements around the sites where the chemical steps of bond cleavage and proton abstraction occur (Figure 11). In particular, the substitution Y116N allows more space around the carbonyl group of Phe64(61) and another water molecule occupies the newly generated space (Figure 11). These changes make a more hydrophilic environment around the carboxamide leaving group, such that proton transfer to the carboxamide oxygen may be more favorable. Although Y116N mutation creates more space around the carbonyl group, most of residues are structurally identical. However, water W_a moves away 0.9 A from Gln42(38). In ALL Q65H, the distance between W_a and Gln42(38) is 2.2 A, which is an unusually short distance. Due to this shift, the same water in LEA is 2.7 A away from Gln42(38). Residue 104 is located between Chain A & B interface and the residue 105 is close to a carbonyl of Gly62 of the chromophore. The residue 104 mutation is Thr to Arg. Because of the bigger side chain between the interfaces, the main chain shift to inside the β-barrel (the difference is 0.49 Å, based on C_α position of 104/105). Also due to the shift, Asn105(LEA) is closer to the carbonyl of Gly62

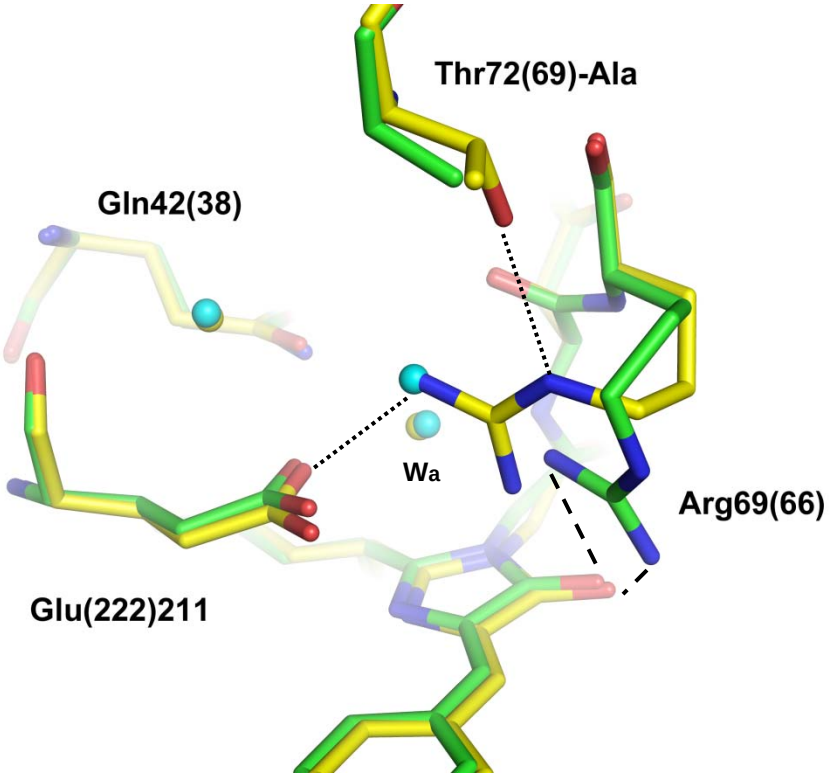


Figure 10. The Thr72-Ala mutation changes the conformation of Arg69. In ALL Q65H (carbon atoms shown in yellow), Thr72 makes a hydrogen bond with Arg69. In LEA (shown in green), no interaction between the Ala72 and Arg69 side chains is observed.

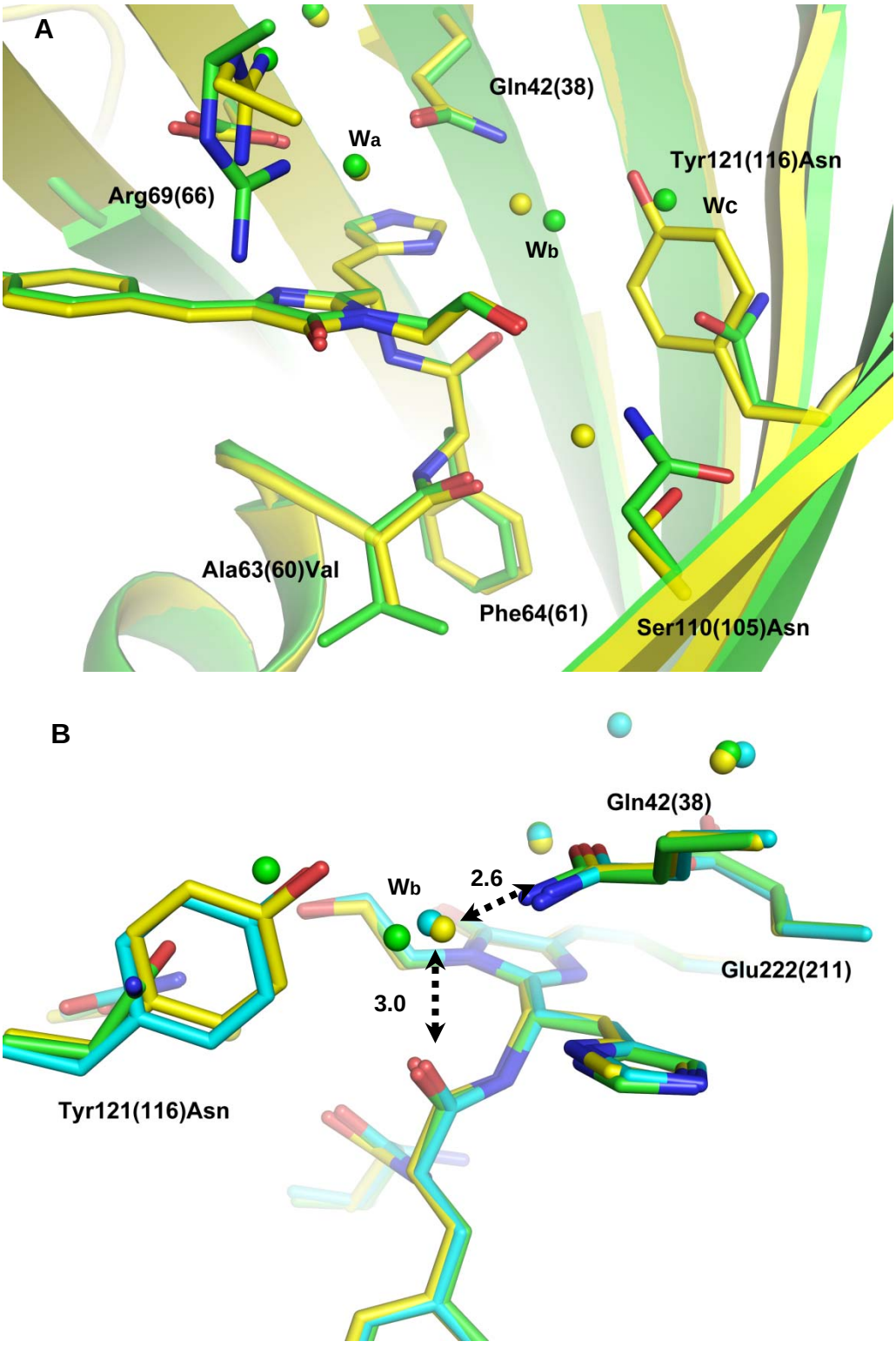


Figure 11. A newly created space by Tyr121-Asn mutation (A) ALL-Q65H (yellow) and LEA (green) (B) ALL-Q65H (yellow), LEA (green), and LEA X121 (cyan); The distances (Angstrom) are based on LEA structure.

and Asn116. These changes may optimize the hydrogen bond network around the chromophore, especially the carbonyl group of Phe64(61).

Quaternary structure and the solvent access channel to the chromophore

When the structure of any monomer of ALL-Q65H (or ALL-GFP) was compared with LEA, I found some interesting differences between two structures. Although the overall structures between two are almost identical, the C-terminus of LEA lacks electron density information (Figure 12). This indicates that the C-terminal of photoconversion protein is more disordered than that of non-photoconversion FP. Interestingly, among 13 mutations, four of them, Arg194(204)-Cys, Arg216(227)-His, Tyr217(DEL), and Met219(228)-Gly, are located in this region (Figure 6). Therefore, the mutations in this region appear to cause increased flexibility of the C-terminus of LEA. In ALL-GFP and ALL Q65H structures, side chains of Glu196(206) and Arg194(204) make salt bridges with the C-terminal chain from the adjacent monomer (Figure 13); ALLGFP (Arg194 – Leu220 / Glu196 – Thr223) and ALL Q65H (Arg194 - Leu220 / Glu196 - Gln223 & Ala224). However, R204C mutation eliminates these interactions, thus the C-terminus of LEA is more flexible. Other 3 mutations also contribute the disorder of C-terminus. Along with the disordered C-terminal chain, I have made another intriguing observation around the C-terminal region, which is a solvent accessible channel that leads from the protein surface to the chromophore's phenolic end (Figure 14 and 15). The solvent access channel in GFP-like proteins has previously been observed in TurboGFP (*77*), zRFP574 (*78*), mKate (*79*), & Dronpa (*80*). The pore is located at the end of a cleft-like structure between β7 and β10 strands and

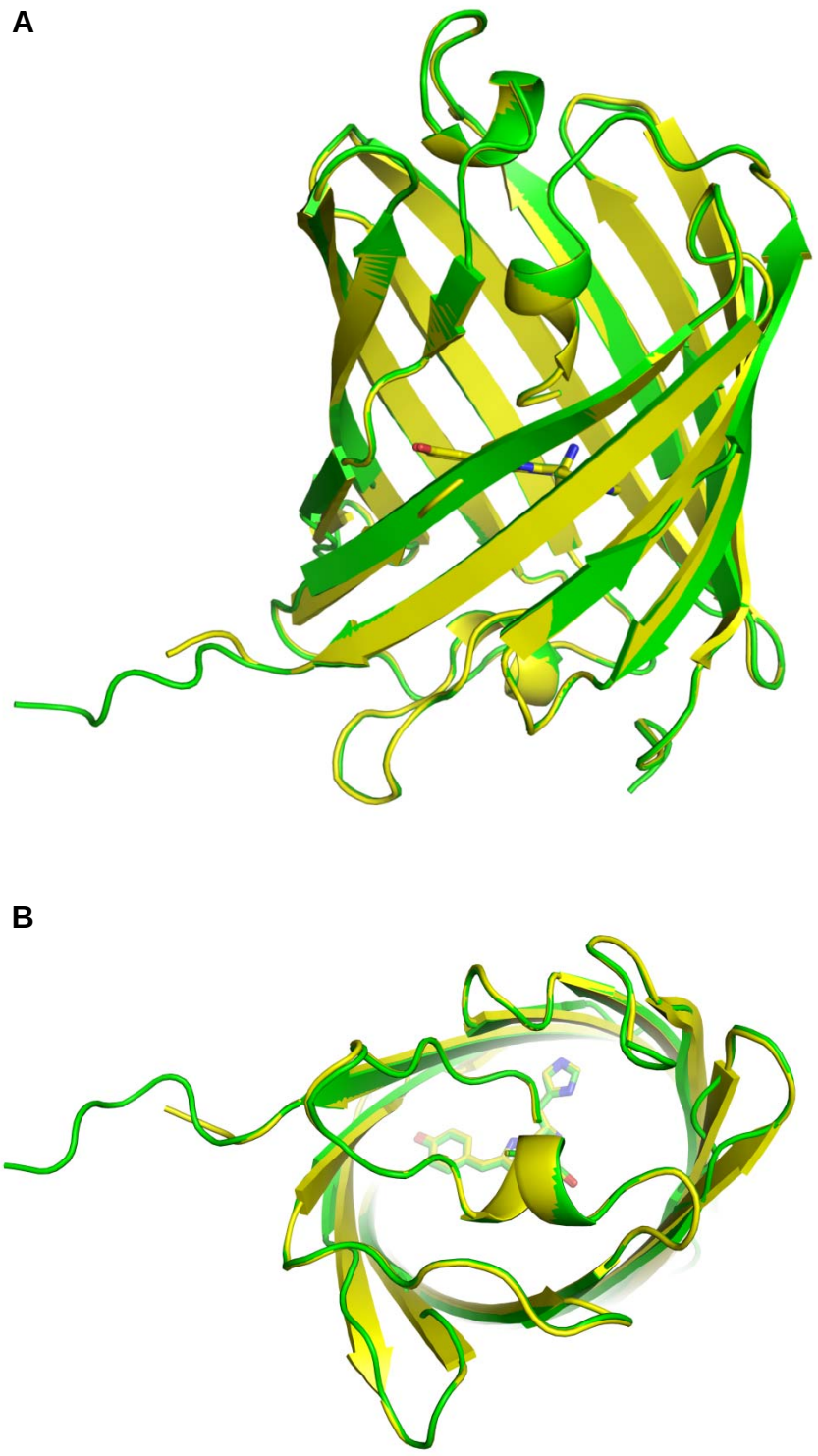


Figure 12. Structural comparison between ALL Q65H (green) and LEA (yellow).

formed by the by the backbone of β7 (Trp139-Glu140-Pro141) and β10 (Arg194-Ile195-Glu196) strands (Figure 15). The Trp-Glu-Pro motif is largely conserved in GFP-like proteins from *Cnidaria* species. Although the channel is not a general feature among GFP-like proteins, it has been suggested that it might be essential for faster chromophore maturation, providing access for molecular oxygen. Both non-photoconvertible and photoconvertible FPs seem to have the solvent channel. However, there is a distinct difference between ALL-GFP / ALL Q65H and LEA variants around the entrance of the solvent channel. As stated earlier, the C-terminal chain interacts with the adjacent monomer. Because of these interactions, the C-terminal end covers the pore in ALLGFP & ALLQ65H structure. In contrast, the entrance of the pore is open to bulk solvent in LEA, because of the more disordered the C-terminal chain and the mutation from Arg194(204) to Cys (Figure 13). Therefore, I made a hypothesis that the excited-state deprotonation of the chromophore is coupled to a proton transfer reaction through a water channel that leads to bulk solvent. If this model is correct, blockage of this channel in LEA should reduce or abolish photoconversion. To test the hypothesis, I have introduced the mutation C204R into the LEA. Although the mutant, LEA C204R, exhibits photoconversion activity, the overall photoconversion process is not as efficient as LEA.

Photophysical Properties of ALL-GFP / LEA variants

ALL-GFP is the ancestral protein of suborder *Faviina*. Its chromophore has Gln62-Tyr63-Gly64 tripeptide and displays green fluorescence (*19, 20*). Kaede-type FPs chromophore consists of His, Tyr, and Gly, and makes the photoconversion

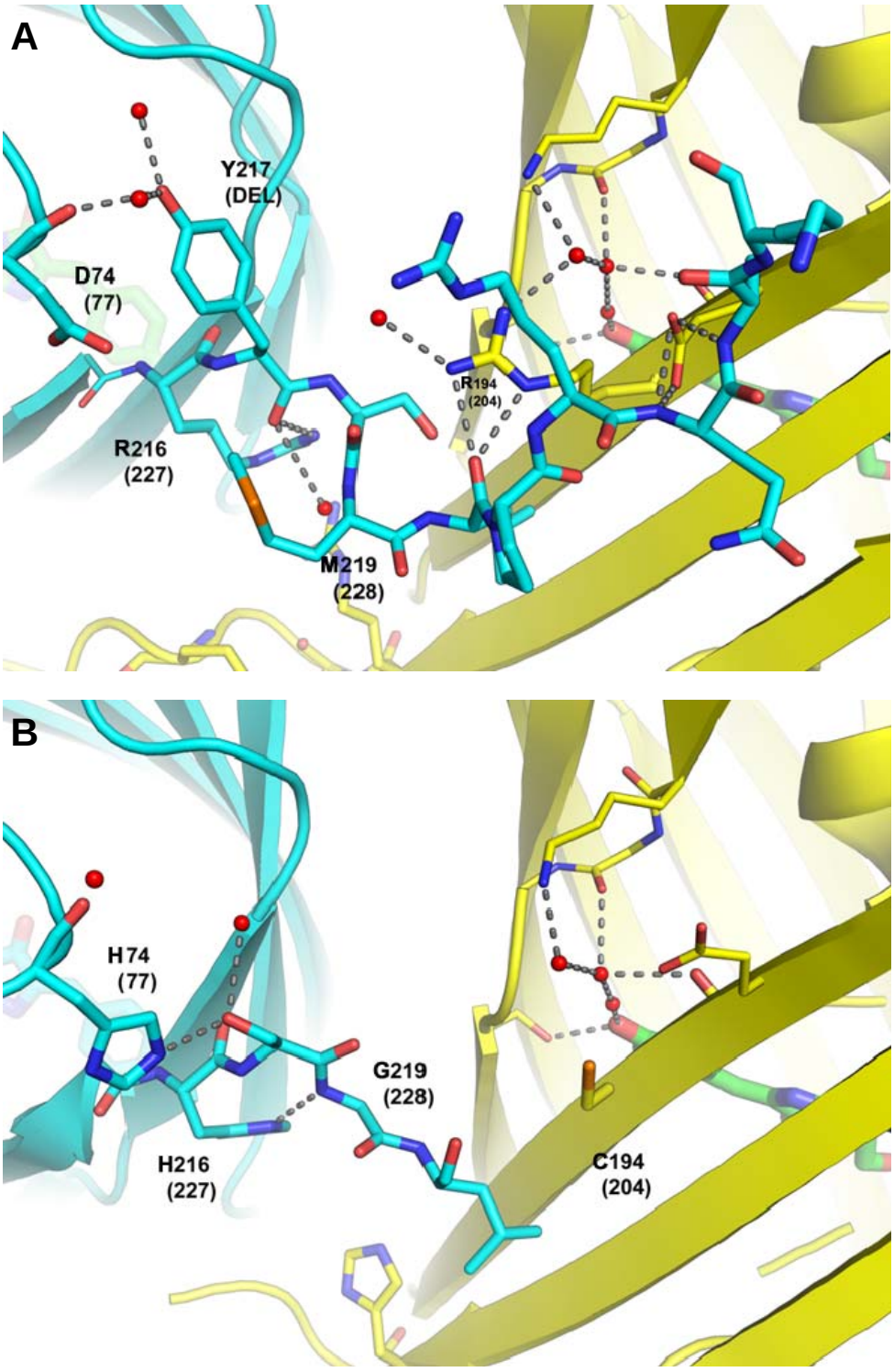


Figure 13. The C-terminal chain of (A) ALL Q65H and (B) LEA. Dashed line represents either hydrogen bonds or salt bridges. Only the residues from LEA mutations are labeled.

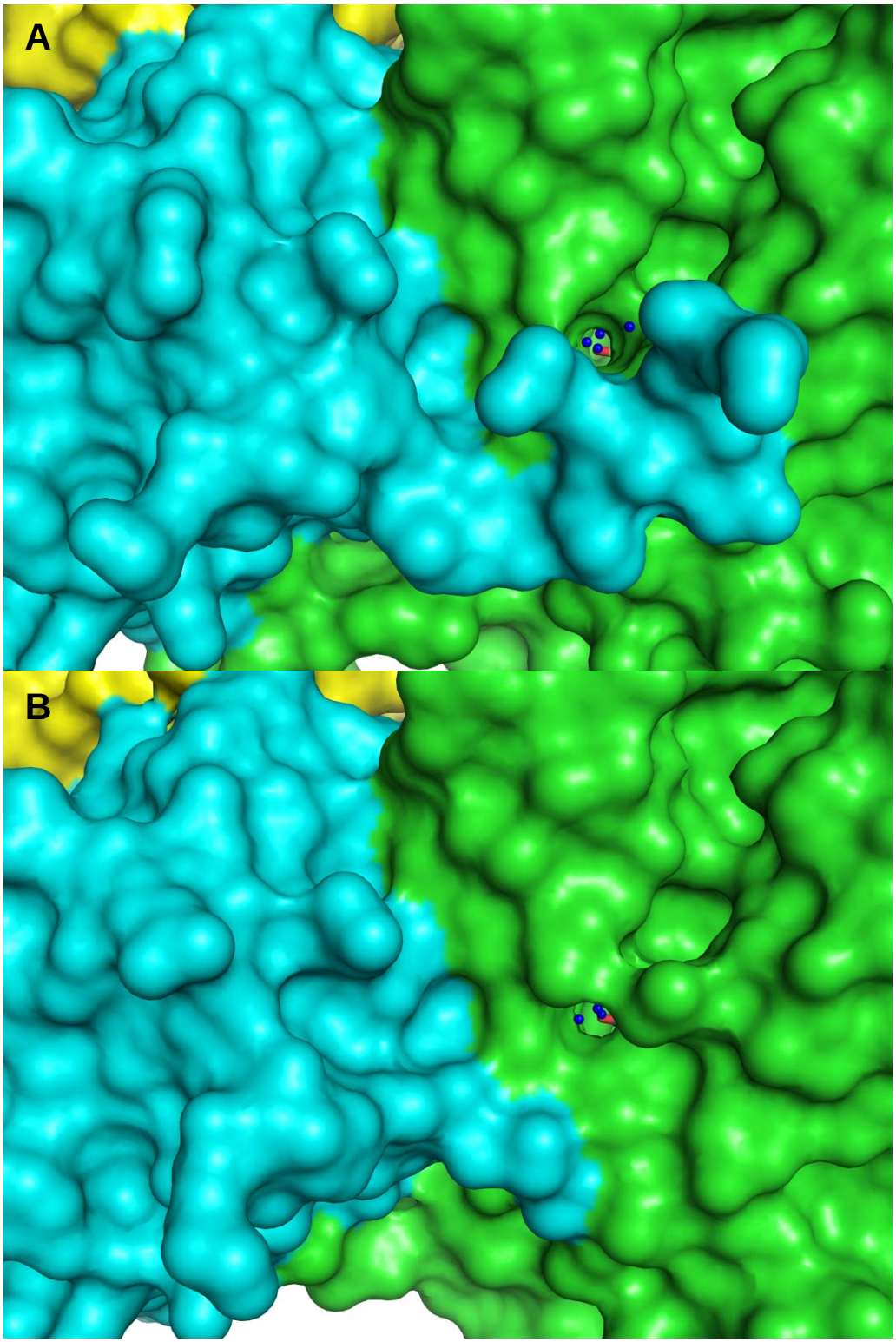


Figure 14. The putative solvent channel – Surface presentation of (A) ALL Q65H and (B) LEA. The chromophore is shown in stick representation and water as blue sphere.

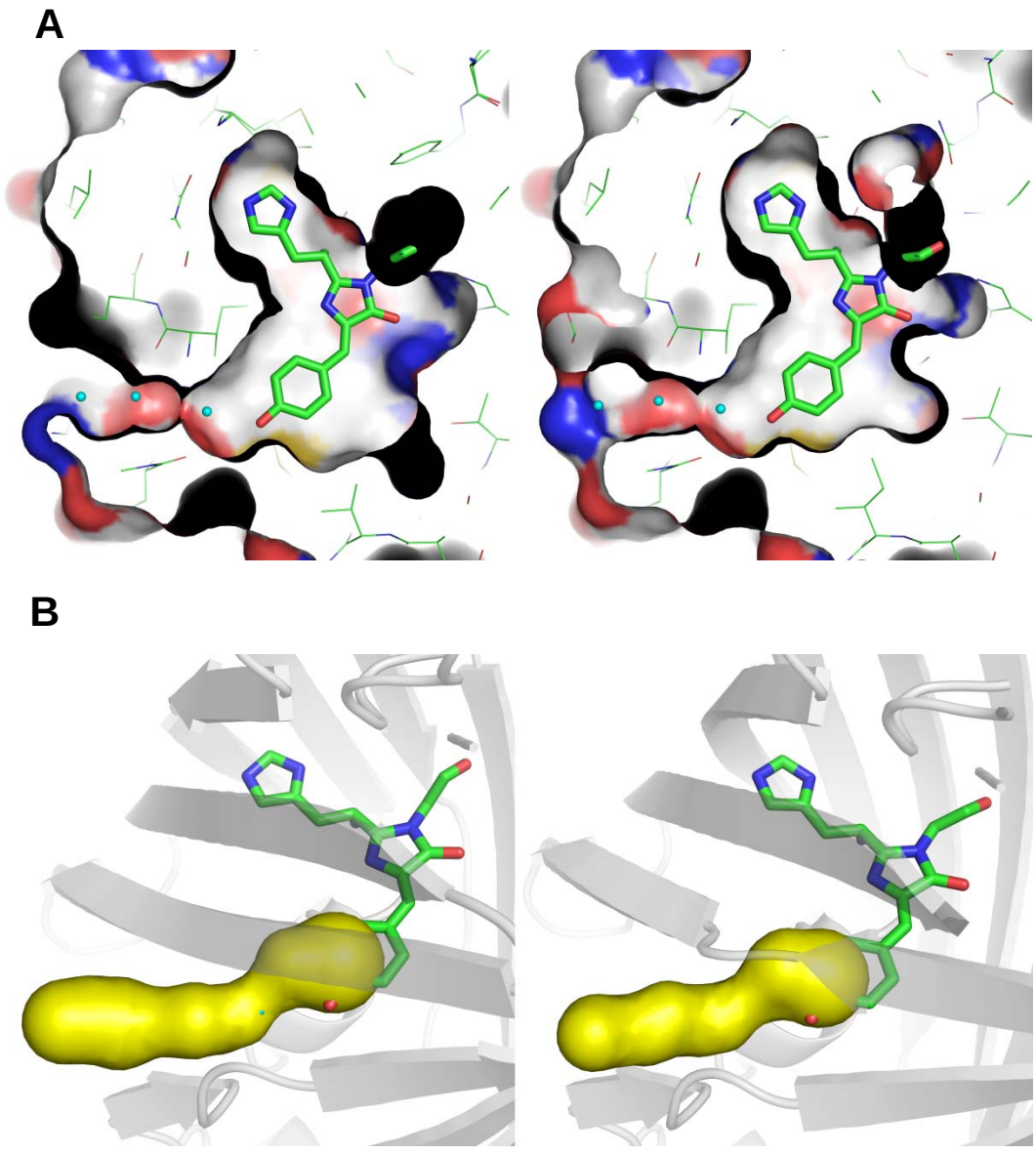


Figure 15. The solvent channels of LEA Q65H (right) and LEA (left). (A) Surface presentations (probe radius 1.4 Å) of the channels and the chromophore cavities. The chromophores shown in stick models and water molecules (cyan) shown in sphere. (B) The putative water channels lead to the chromophore, calculated using the software CAVER (*81*) shown in yellow. The proteins are shown in cartoon presentation and the chromophores shown in stick.

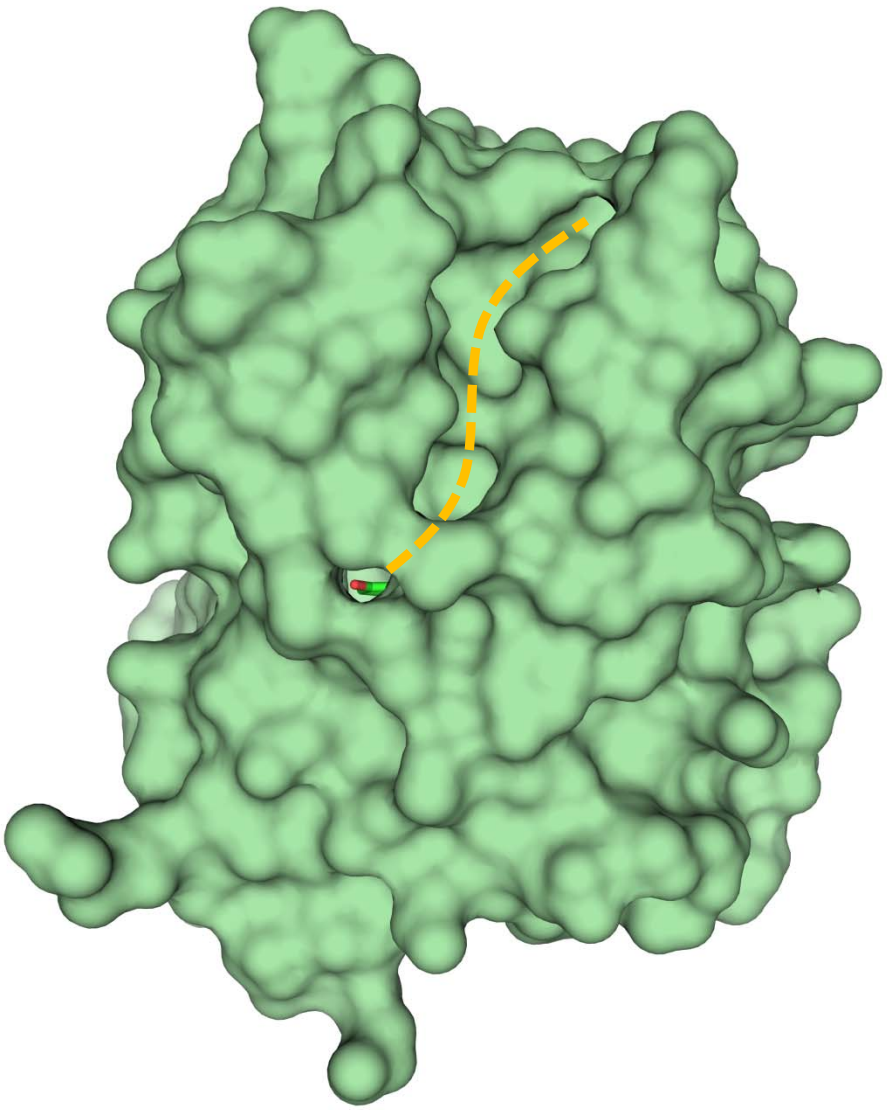


Figure 16. The entrance of solvent channel and the cleft on the surface of LEA. Surface view of LEA showing the solvent channel and the cleft. The chromophore is shown as a stick model. The cleft is highlighted by a dashed line.

reaction with UV light. ALL Q65H also has also the same His-Tyr-Gly chromophore like other Kaede-type FPs. Although the His-Tyr-Gly chromophore is required for the red fluorescence, ALL Q65H does not display the photoconversion reaction and only shows green fluorescence. All LEA variants have His-Tyr-Gly chromophore and also are capable of making the photoconversion reaction. The absorption spectra of green ALL-GFP and LEA variants are shown in Figure 16. All of them displayed a major absorption wavelength maximum around 490-510 nm with a slight shoulder at 480 nm. A relatively minor peak was found around 370-380 nm. At low pH, the peak around 380 nm is increased and the 510 nm peak is decreased, with isosbestic points around 410-420 nm. The isosbestic points show that there is an interconversion between two species. Like other FPs, this absorbance change in different pH values indicates that the 380 nm peak corresponds to the neutral chromophore and the 510 nm does to the anion one. Kaede-type FPs shows single-site protonation equilibrium between the neutral and the anionic forms (*39, 41, 52*). Therefore, the pH dependence of the anionic form follows the Henderson-Hasselbalch relation (Figure 18). The chromophore proton dissociation constant (pK_a) was determined for each variant in the GFP-like pre-photoconversion state (Table 9). The proton dissociation constants at the phenolic end of chromophore were generally low, ranging from 5.2 to 6.2 for ALL-GFP and the LEA variants, as has been found to be the case for Kaede (pK_a = 5.6) (*41*), EosFP (pK_a = 5.8) (*39*). However, LEA-X72 (pK_a = 8.4) and all three LEA Q38 mutants (pK_a = 7.9 and 7.5) display significantly higher values than others. Higher pKa values of the Kaede-type FPs were previously reported in Dendra2 (pK_a = 7.1) (*52*) and KikGRX (pK_a = 8.0) (*42*). In addition LEAX72 shows considerable anti-cooperative titration

Table 9

Properties of ALL-GFP / LEA variants.

Protein	pKa[a]	Hill Coefficient[b]	The Quantum Yield[c]
ALL-GFP	5.20 ± 0.06	0.98	0.78
ALL Q65H	5.50 ± 0.01	1.00	0.79
LEA X6	6.24 ± 0.12	0.67	0.77
LEA	6.34 ± 0.14	0.60	0.81
LEA X72	8.43 ± 0.02	0.49	0.73
LEA X121	5.92 ± 0.05	0.68	0.88
LEA C204R	5.73 ± 0.06	1.62	-
LEA Q38A	7.79 ± 0.06	0.62	-
LEA Q38M	7.79 ± 0.04	0.67	-
LEA Q38N	7.51 ± 0.06	0.76	-

a and b : pKa and Hill coefficient were calculated based on the Henderson-Hasselbalch equation. See page 33
c: The quantum yields were calculated with using fluorescein as a standard. See page 34.

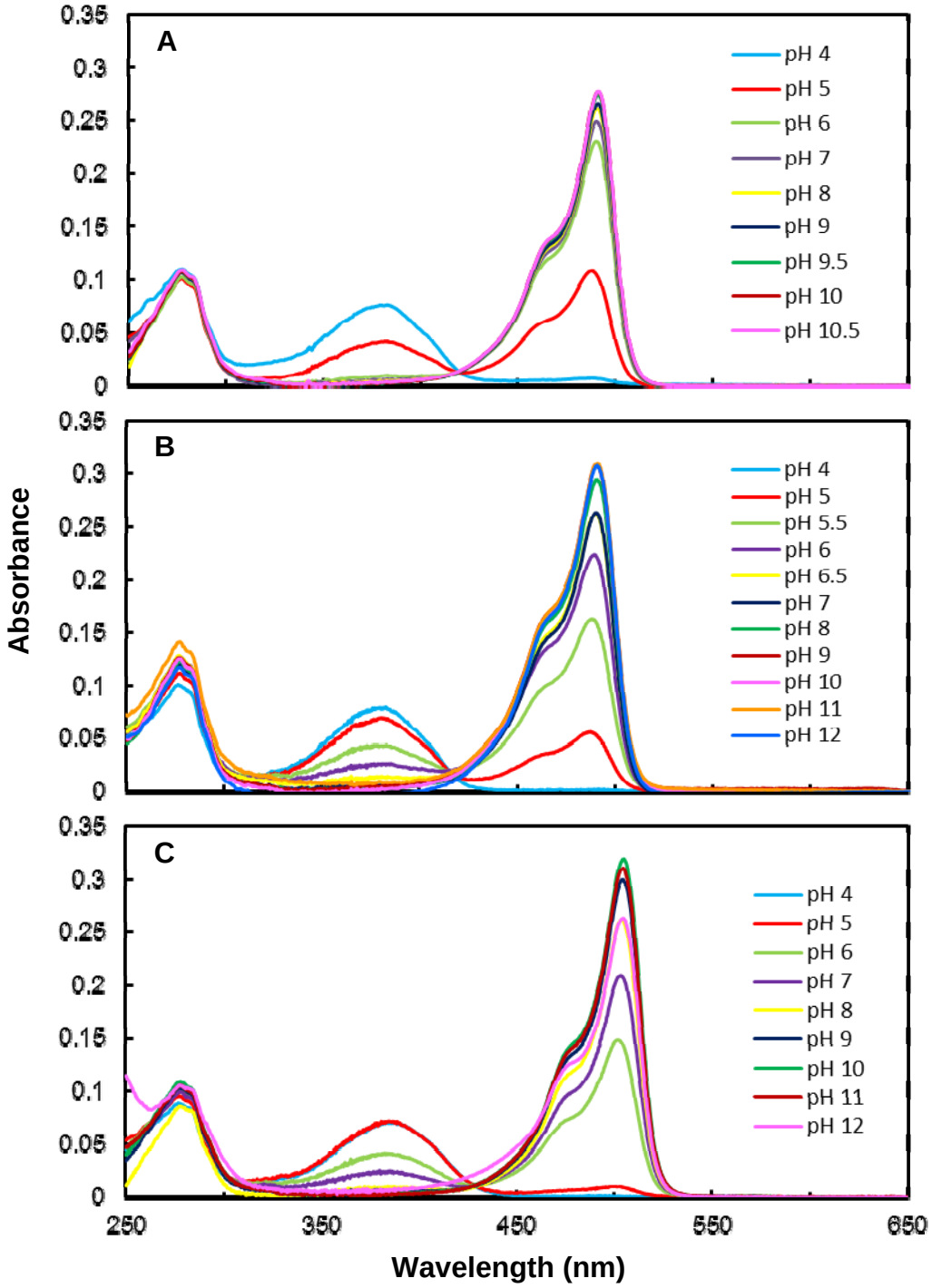


Figure 17. Absorbance spectra of green ALL-GFP/ LEA variants at different pH values – (A) ALL-GFP, (B) ALL Q65H, (C) LEA, (D) LEAX(6), (E) LEA X72, and (F) LEA X121

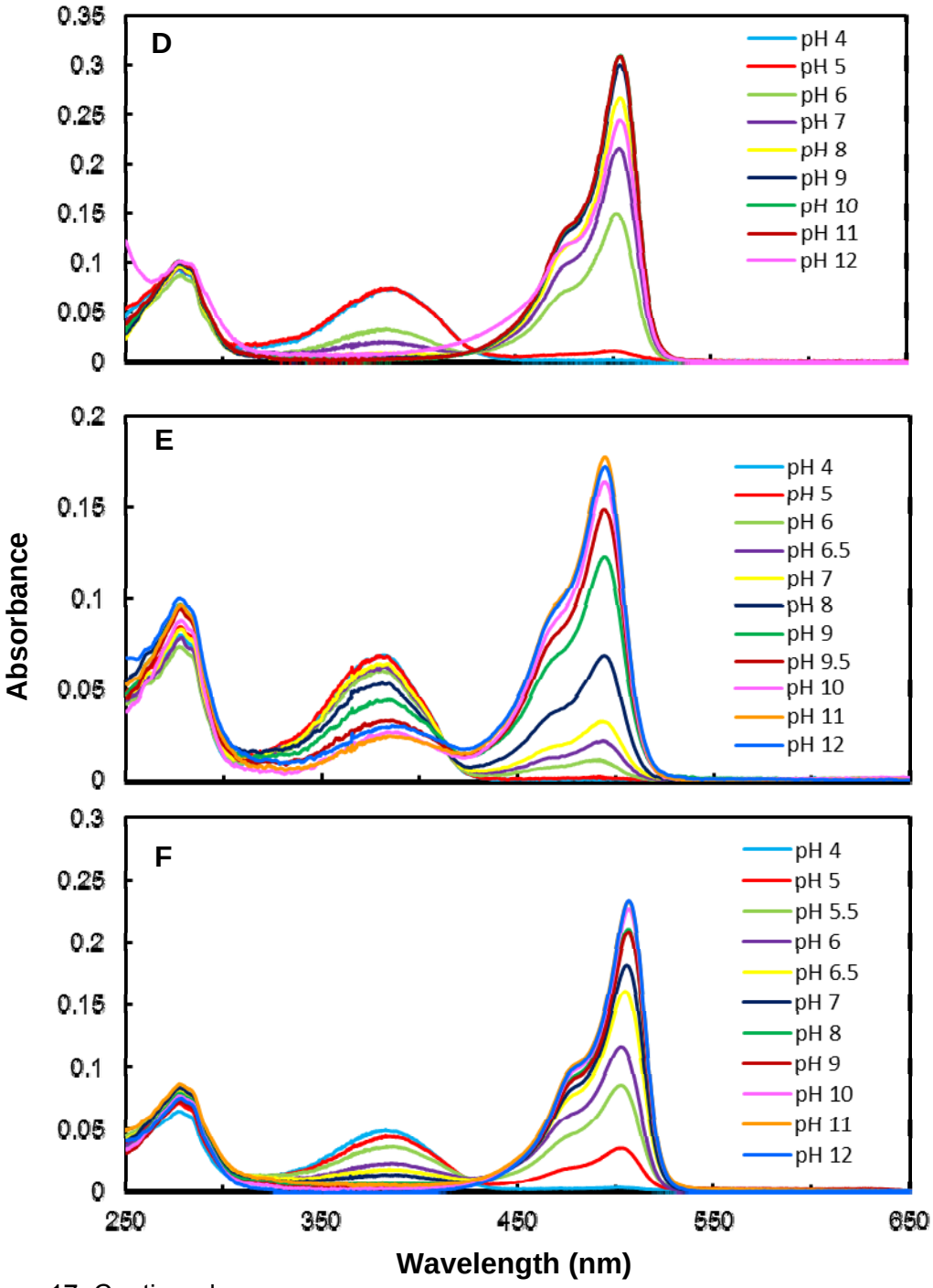


Figure 17. Continued

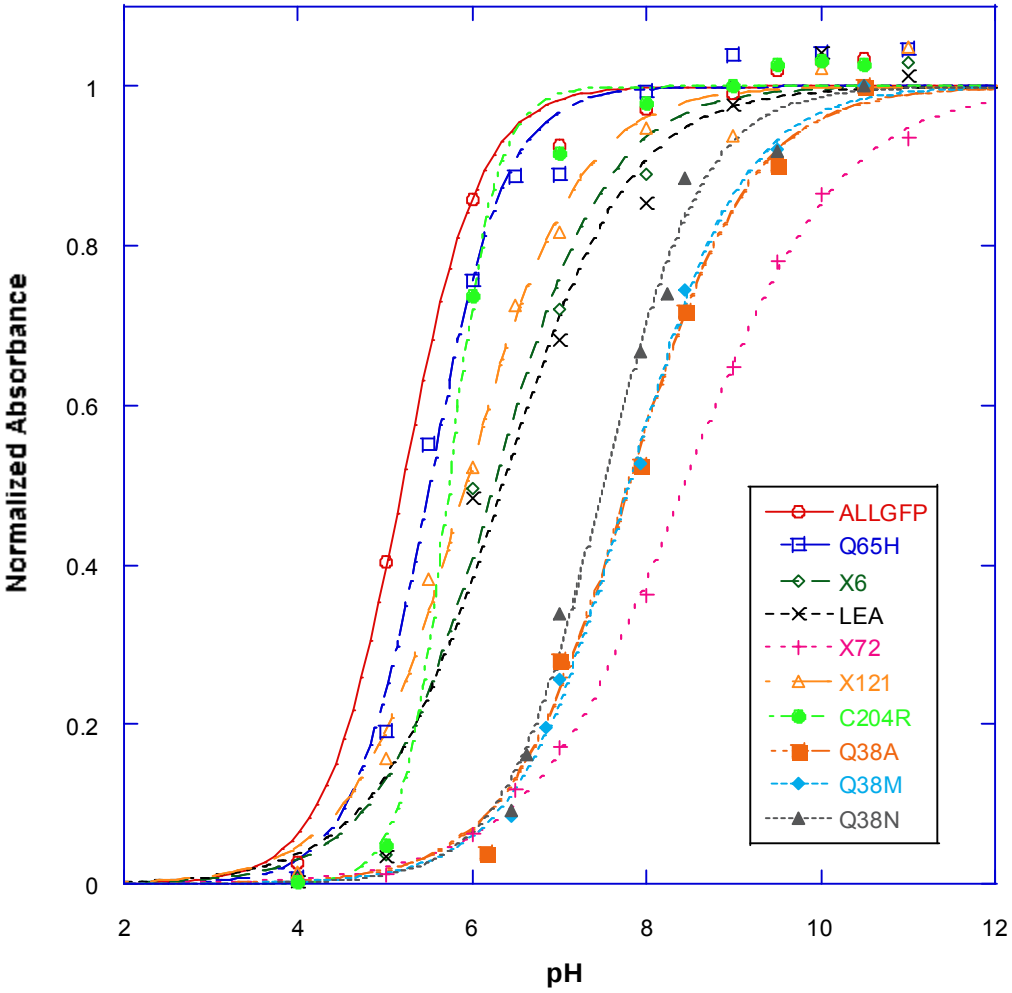


Figure 18. pH dependencies of the anionic green chromophores of ALL-GFP and LEA variants. An absorbance peaks of anionic green chromophore at various pH values were normalized (ALLGFP: 490 nm, ALL Q65H: 490 nm, LEA X6: 503 nm, LEA: 504 nm, LEA X72: 494 nm, LEAX121: 506 nm, LEA C204R:, LEA Q38A: 507 nm, LEA Q38M: 507 nm, LEA Q38N: 507 nm). For the pKa values for each protein, see Table 9.

behavior, as indicated by a Hill coefficient n of 0.5. Other LEA variants exhibited anti-cooperativity to a lesser degree, with n ranging from 0.6 to 0.7, whereas ALL-GFP an ALL-Q65H did not exhibit any cooperativity at all (Table 6).

Excitation and emission spectra of the ALL-GFP / LEA variants were collected from pH 4.0 to 12 (Figure 19). The green fluorescence emission of ALL-GFP peaks at 502 nm, with a quantum yield of 0.78. Compared to ALL-GFP and ALL Q65H, both excitation and emission peaks are red-shifted in LEA variants (Table 10). However, the emission spectrum LEA-X72 shifted toward red by 15 nm (Figure 20). This may come from the different side chain conformation of Arg69(66), as described in the Dendra2 study. Because of Thr72(69)-Ala mutation, in the Dendra2 structure, the hydrogen bond between Arg69(66) and the carbonyl oxygen of the imidazolinone ring is missing (Figure 9) (*52*). Due to the lack of this interaction, the anionic form is less favorable than the neutral form and LEA-X72 shows a higher pKa value. Also this lack of the interaction between Arg69(66) and the chromophore results in a blue shift of the spectra. The pH dependence of the emission spectra follows the dependence of the absorbance. With increasing pH, emission by the anionic chromophore gets larger, as the neutral chromophores decreases. UV irradiation (405 nm) causes the green-to-red photoconversion. The maximum of the red absorption is located at 572 nm. Also 533, 460, and 340 nm peaks appeared. These peaks also reflect the neutral (460 nm) and anionic (572 nm) state of the chromophore (Figure 21).

Table 10

Fluorescent properties of ALL-GFP and LEA variants.

Protein	Green fluorescence[a]		Red fluorescence[b]	
	Excitation (nm)	Emission (nm)	Excitation (nm)	Emission (nm)
ALL-GFP	489	502	-	-
ALL Q65H	490	502	-	-
LEAX(6)	502	513	569	578
LEA	503	513	571	580
LEAX121	506	514	571	579
LEAX72	491	505	568	582

a: See Figure 18 and 19 for the spectra
b: see Figure 20 for the spectra

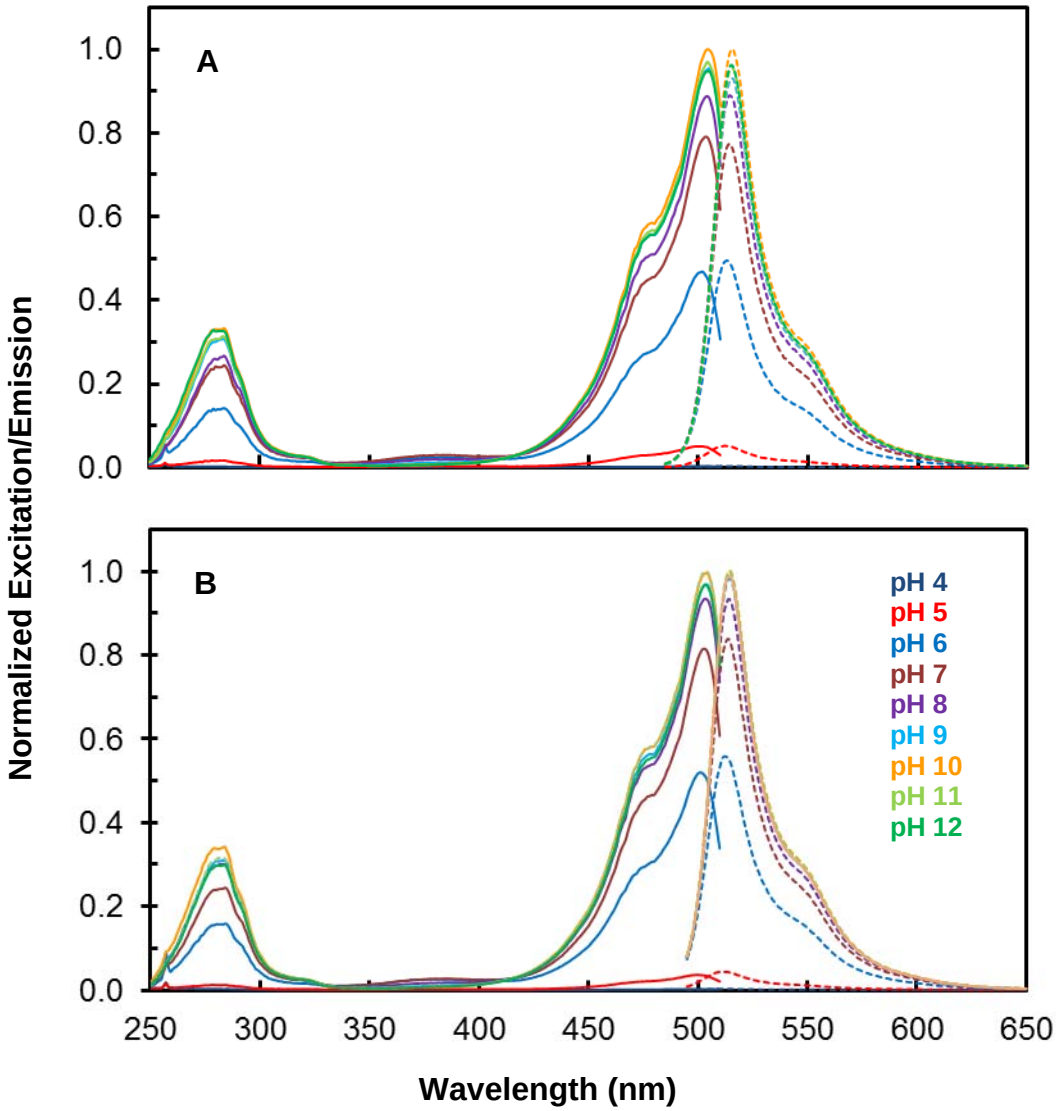


Figure 19. Fluorescent spectra of green (A) LEA and (B) LEAX(6) at different pH values – excitation (solid line) and emission (dotted line)

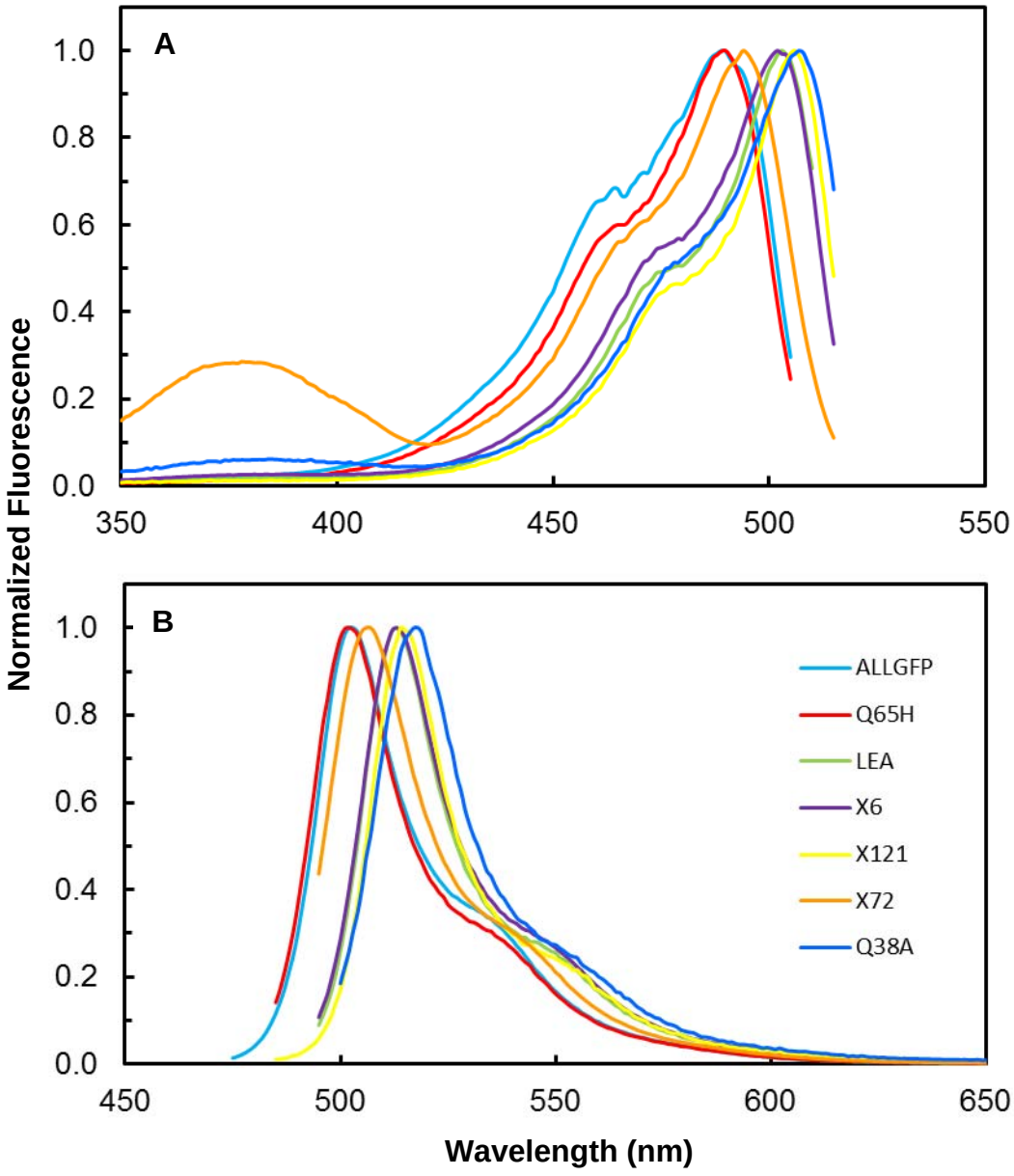


Figure 20. Fluorescence spectra of ALL-GFP / LEA variants – (A) excitation and (B) emission spectra

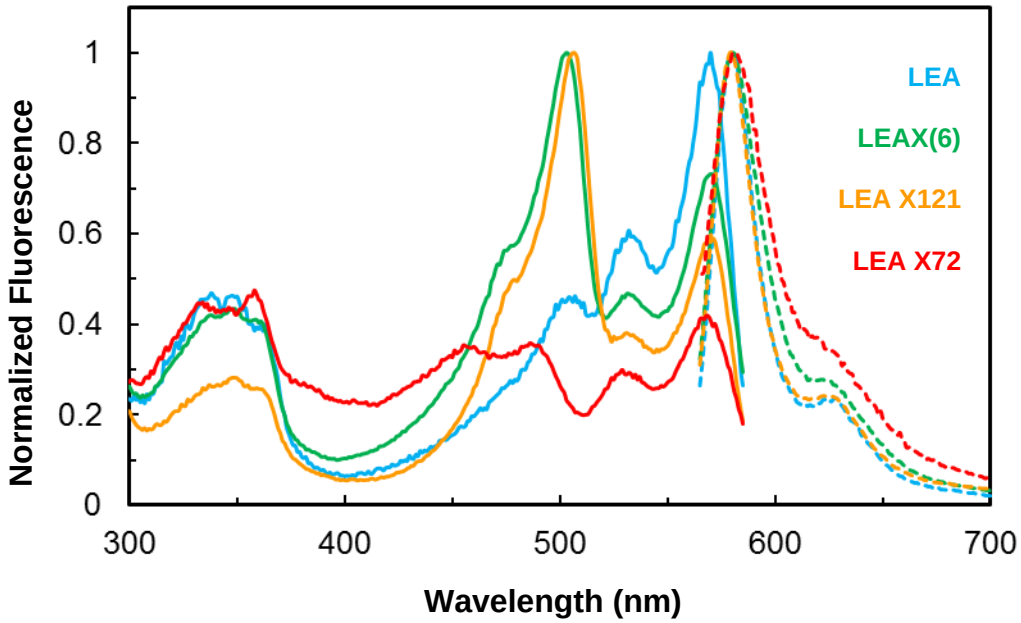


Figure 21. Fluorescence spectra of the red LEA variants (at pH 7.9) – excitation (solid line) and emission (dotted line)

Green-to-red photoconversion of LEA variants

The green LEA is efficiently converted to the red LEA by illumination with 405 nm light. The proposed hypothesis suggests that photoconversion starts from the neutral green chromophore. Because of this, the photoconversion efficiency would be proportional to the concentration of neutral chromophore and the previous research agreed with the hypothesis (*39, 41, 52*). To examine the efficiency of photoconversion reaction among LEA variants, 0.1 mg/ml of green protein solutions were illuminated with 405 nm UV light and the absorbance of samples were measured from 250 – 650 nm. Initially, photoconversion reactions were carried out at pH 7.9, which is the protein stock solution buffer (Figure 22). As the reaction continues, the peaks at 500 and 480 nm decreases and the 530/570 nm peaks increases. At pH 7.9, LEA showed the fastest reaction. Next fastest one is LEA X(6) and followed by LEA X121. LEA X72 is the slowest. In addition, ALL Q65H did not make the photoconversion reaction at all. LEA X(6) has all of 'large effect' mutation residues and lacks four 'fine-tuning' mutations. Therefore, the reaction rate is relatively comparable to LEA. LEA X72 and LEA X121 has reversion to ALLGFP, which are A72(69)T and N121(116)Y, respectively and both are one of the 'large effect' residues (*62*).

Next, to verify pH dependence of photoconversion reaction, 0.1 mg/ml of LEA was illuminated with 405 nm UV light at 13 different pH values for 60 min. During the reaction, the absorbance spectra were intermittently measured from 250 – 650 nm (0, 1, 2, 3, 5, 7, 10, 15, 20, 30, 45, and 60 min). In Figure 23, absorbance spectra during photoconversion reaction at 3 different pH values are displayed.

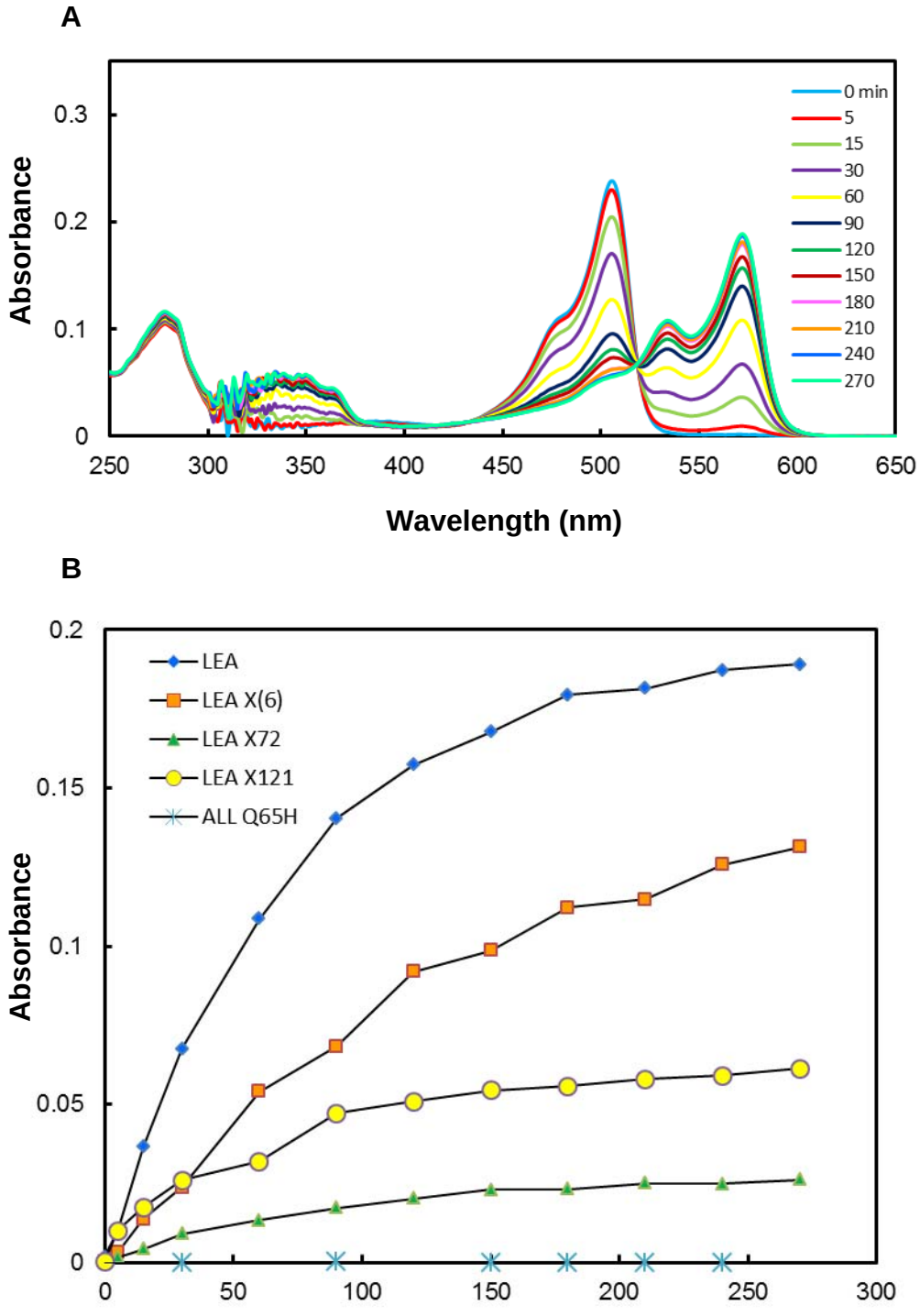


Figure 22. Photoconversion reaction of LEA variants at pH 7.9. (A) The absorbance spectra of LEA Photoconversion (B) The absorbance change of anionic red chromophore during photoconversion reaction.

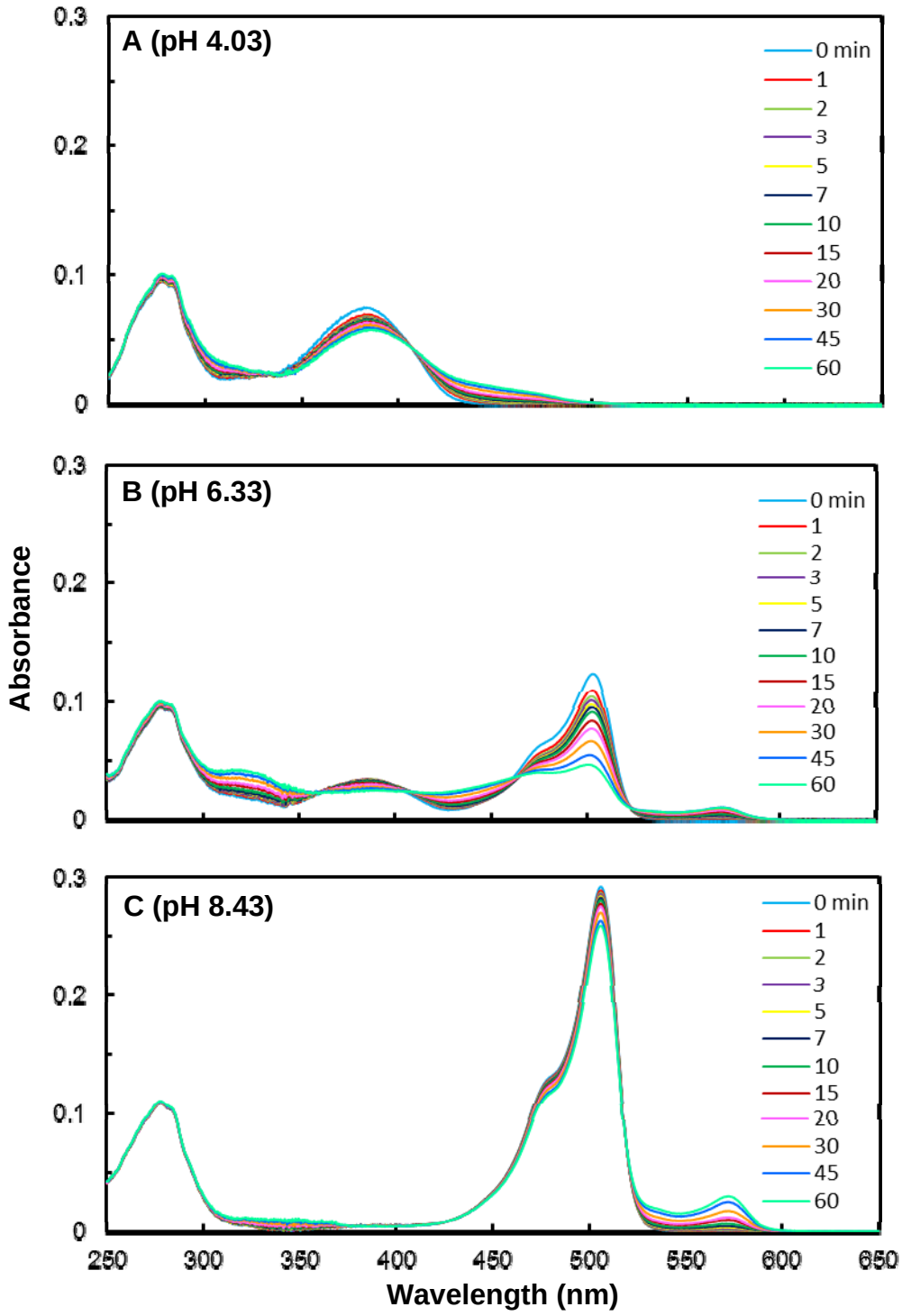


Figure 23. Photoconversion reactions of LEA at different pH values.

At pH 4.03, most of the chromophore is in the neutral form. A peak at 380 nm is converted to a 460 nm peak, which is a neutral red chromophore. At pH 6.33, pK_a of LEA, both neutral (380 nm) and anion (500 nm) chromophore peaks disappears and the red neutral (460 nm) and anion (570 nm) peaks increases over the reaction. Lastly, at high pH (at pH 8.43), the chromophore only exists in the anionic form. As the reaction goes, the green anionic peak (570 nm) decreases, and the red anionic peak (570 nm) increases. To describe the kinetics of photoconversion reaction, total amount of green chromophore was calculated from absorbance of the neutral and anion chromophores and then plotted against time (Figure 24). Initially, the overall plot looked like a first-order reaction. But neither the first-order nor the second-order rate law applied to the plot, due to an initial high velocity of decreasing green chromophore (from 0 to 1 min), especially in low pH values (Figure 24). Therefore, after excluding the initial fast phase (0 and 1 minute), the plot was fitted to first-order kinetics (Figure 25). The observed rate constants (k_{obs}) versus pH data can be fitted to the bell-shaped equation that is associated with a diprotic system where the two successive pKa values are closer than 3.5 pH units (Figure 26) (*74*). I found pKa_1 is 4.33 and pKa_2 is 7.68, which the difference, 3.30, is less than 3.5 units. The current hypothesis has two important proton transfer steps – His65(62) C_β proton abstraction by Glu211 and a proton transfer from the imidazole group of His65(62) to the carbonyl group of Phe64(61) (Scheme 2). pKa_1 4.33 is similar to a typical pKa value (4.1) of glutamate side chain. The pKa of histidine side chain is 6.0 in aqueous solution. In LEA structure, the imidazole ring of His65(62) is located in hydrophobic environment (Figure 27). Therefore, the pKa would be higher than in aqueous solution

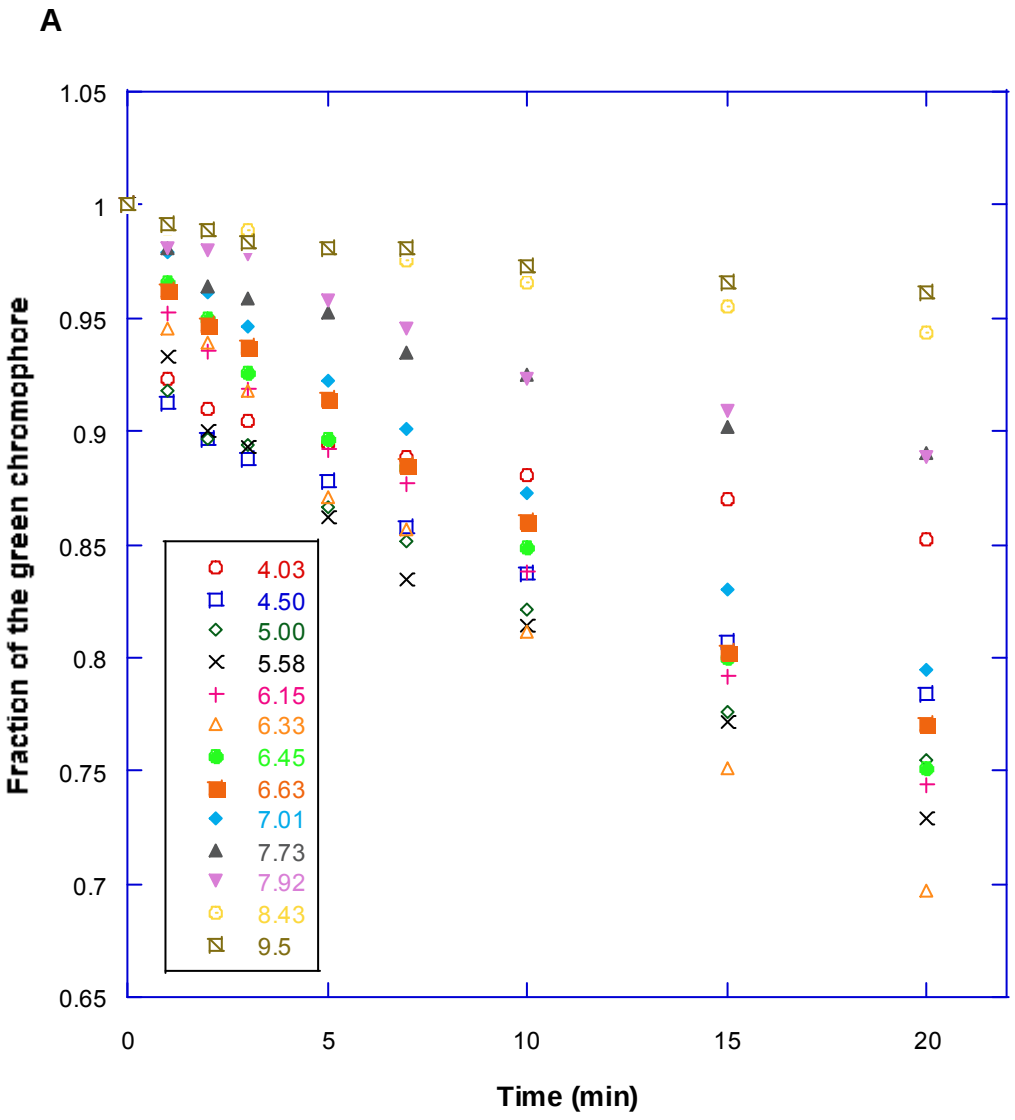


Figure 24. Time course of photoconversion reaction for LEA at various pH values. The photoconversion reaction is described as the change of the green chromophore as time

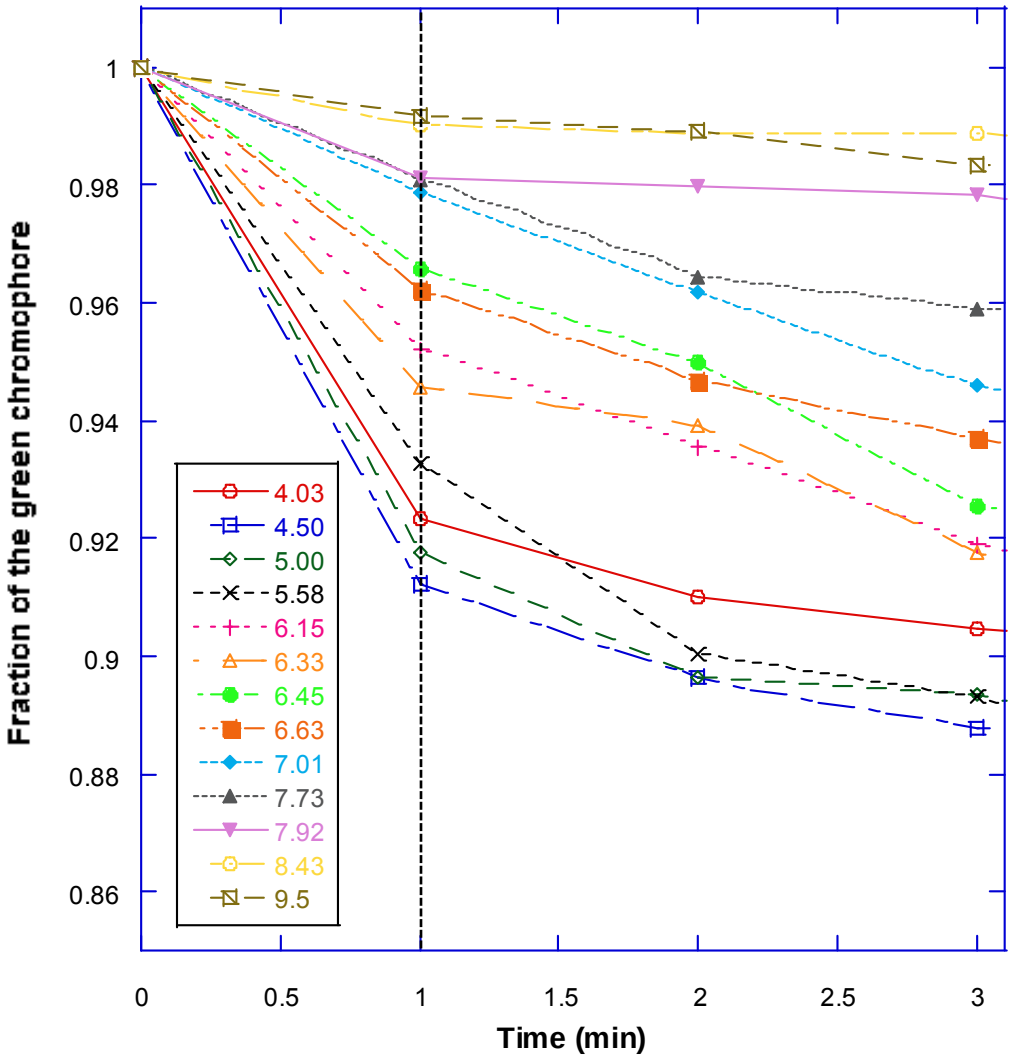


Figure 24. continued.

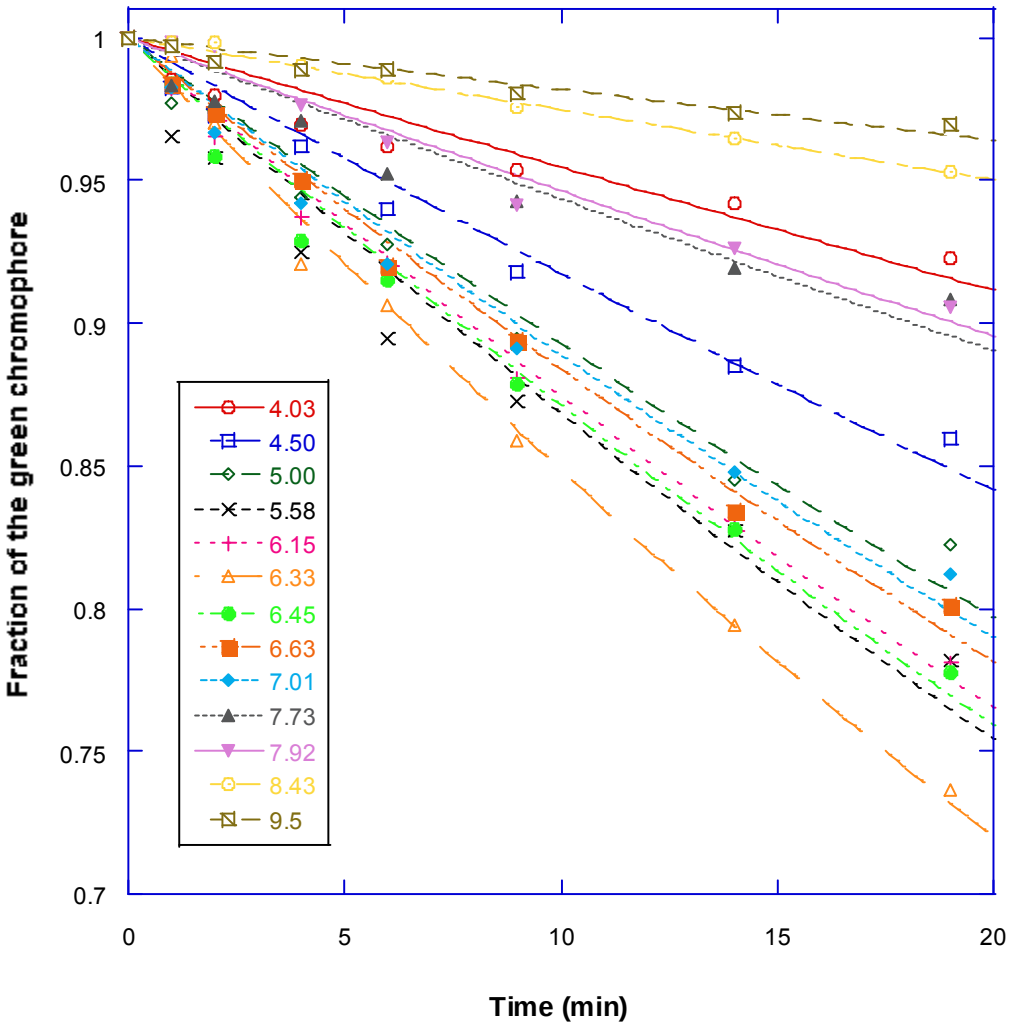
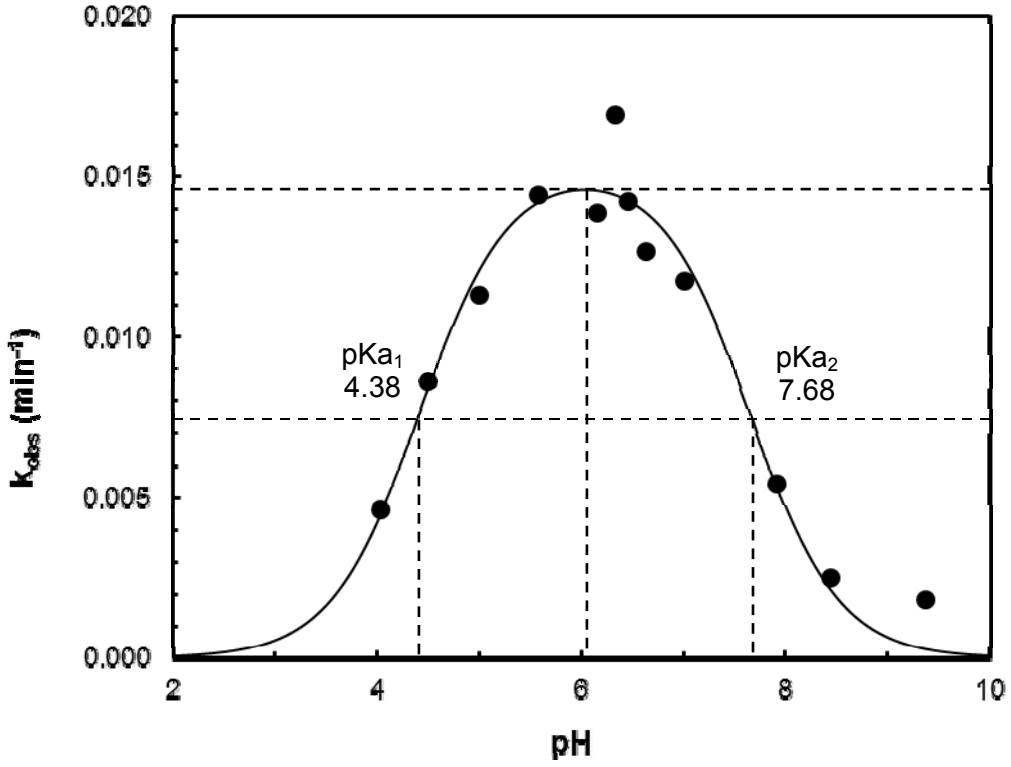


Figure 25. The adjusted time course of photoconversion reactions for LEA. The photoconversion reaction is described as the change of the green chromophore as time. 1 minute time point sets as a starting point.

A



B

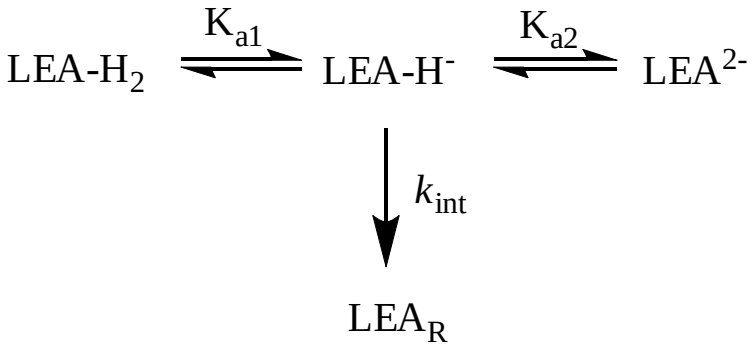


Figure 26. pH dependence of photoconversion reaction. (A) The observed first order rate constant of photoconversion (symbols) and the expected pH dependence (dotted line) derived from a simple diprotic system which the two successive acid dissociation constants. (B) A diprotic system of the photoconversion reaction.

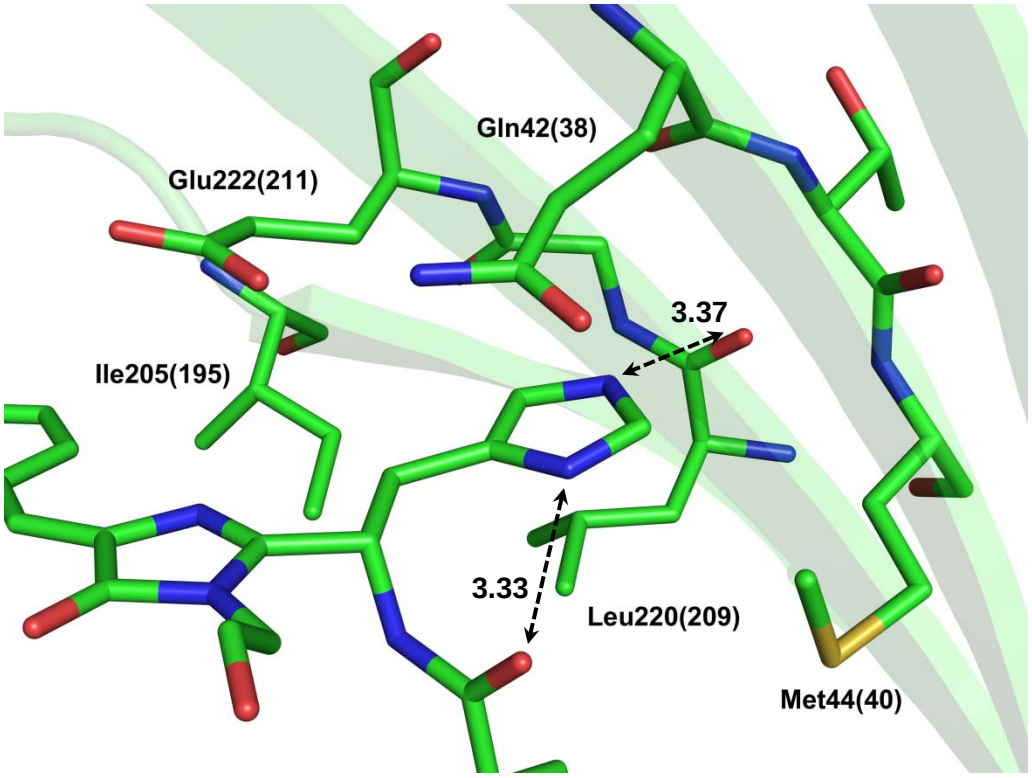


Figure 27. The environment around His65(62) from LEA chromophore (distances are angstrom)

pH dependence of photoconversion reaction of LEA C204R and LEA X72 were also tested. Because photoconversion reaction is started from the neutral green chromophore, pKa value of the chromophore is very important. Therefore, those two variants were chosen, based on their pKa values of the chromophore (LEA C204R pK_a = 5.7 and LEA X72 pK_a = 8.4). If only the pK_a value matters to photoconversion reaction, a protein, which has the highest pKa value, will show the fastest reaction rate. Both proteins were illuminated with 405 nm UV for 10 min. The change of the green chromophore during the reaction is shown in Figure 28. Although both proteins also displayed a two-phase kinetics, neither the first-order rate law nor the second-order rate law applied to the plot, after excluding the initial fast phase. Compared to the fraction change of the green chromophore, LEA C204R and LEA X72 showed a slower reaction than LEA. LEA C204R has a pK_a value 5.7, which is similar to the non-photoconvertible FPs, ALL-GFP and ALL Q65H. As discussed earlier, C204R mutation was introduced to test that residue 204 could block the water channel or not. Although I could not find any evidences to prove the hypothesis, it certainly changes pK_a value of the chromophore. The lower pK_a value means the smaller neutral chromophore fraction. Therefore, photoconversion reaction is slower than LEA. The pK_a value of LEA X72 is 8.4, which is higher than LEA. Based on the pK_a values, LEA X72 is favorable to photoconversion reaction, which generates more neutral chromophore. However, LEA X72 also displayed slower photoconversion reaction than LEA. This observation can be explained with A72T mutation. ALL-GFP and ALL Q65H have Thr at residue 72, which makes the hydrogen bond with Arg69. Due to the interaction, Arg69 of the non-photoconvertible proteins makes the hydrogen bond with Glu222 (Figure 10). In the current hypothesis,

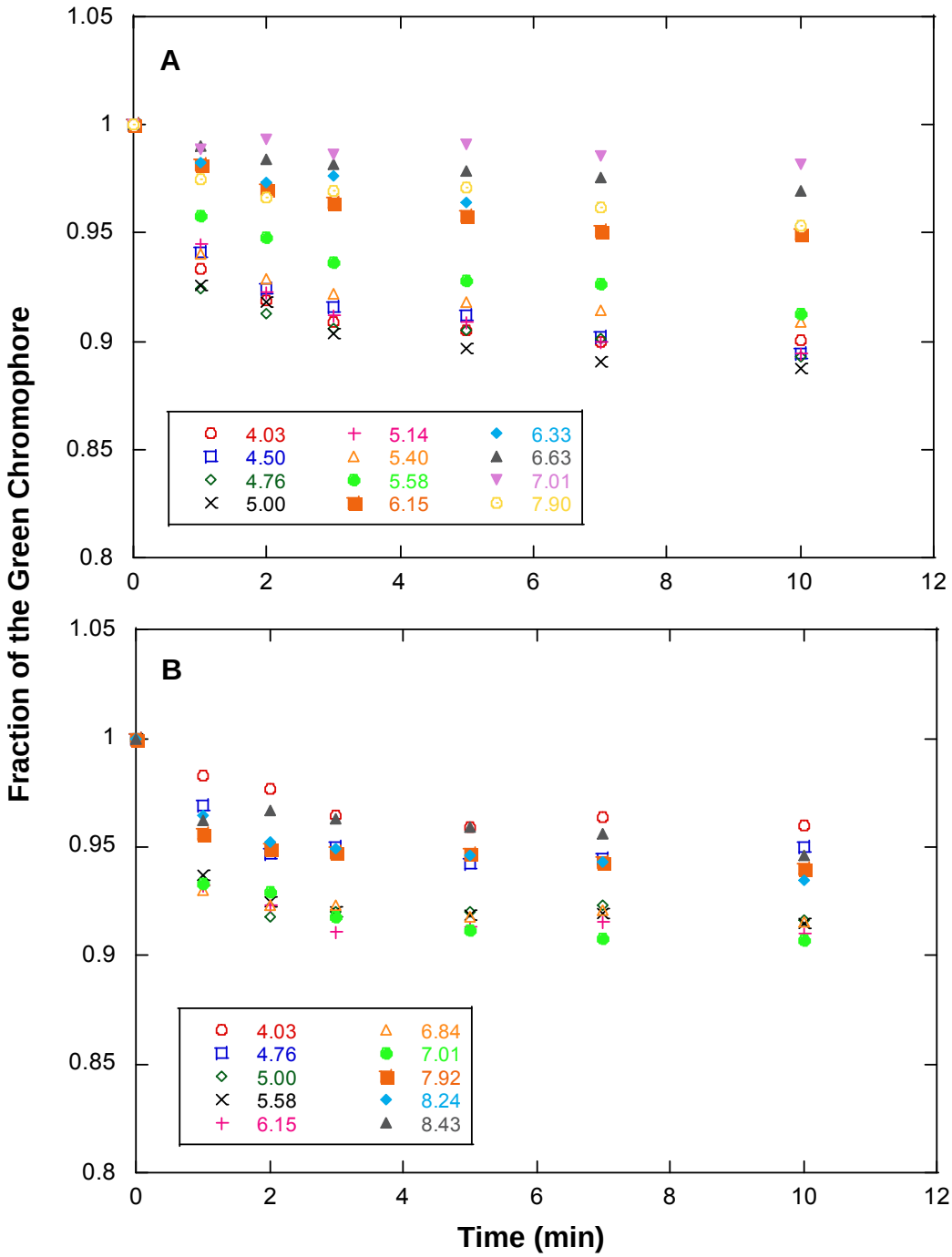


Figure 28. Time course of photoconversion reaction for (A) LEA C204R and (B) LEA X72 at various pH values. The photoconversion reaction is described as the change of the green chromophore as time

Glu222 takes a C_β proton from His65. An additional positive charged residue (Arg69) could make the deprotonation by Glu222 more difficult. Therefore, photoconversion reaction will be slower than LEA.

After photoconversion reaction, a LEA sample was left in the reaction chamber overnight (in the dark). The next day, the color of sample was changed and the absorbance spectrum of sample was measured (Figure 29). Compared to the last absorbance measurement from the reaction, the peak of anionic chromophore was increased and the neutral chromophore peak was reduced in size. To verify this unexpected observation, other LEA variants (LEA X6, LEA X121, and LEA C204R) were also tested. Both LEA X121 and LEA C204R showed similar changes between the neutral and anionic chromophores. However, there is no interchange in LEA X(6). This kind of change is found in photoswitching FPs such as Dronpa (*30*), Pardron (*82*), asFP595 (*33*) and IrisFP (*43*). Dronpa, Padron, and asFP595 are a reversible photoswitching green fluorescent protein. In IrisFP, a photoswitching / photoconversion FP, both green and red chromophores exhibit a reversible photoswitching reaction, resulting from a *cis-trans* isomerization of the chromophore (*43*). Since this change happens in the dark, this observation would be a thermal isomerization of the chromophore. Generally, the thermal decay is much slower than the protonation-deprotonation equilibrium process. This type of phenomenon was also reported in other photochromic FPs (*31, 80, 82, 83*).

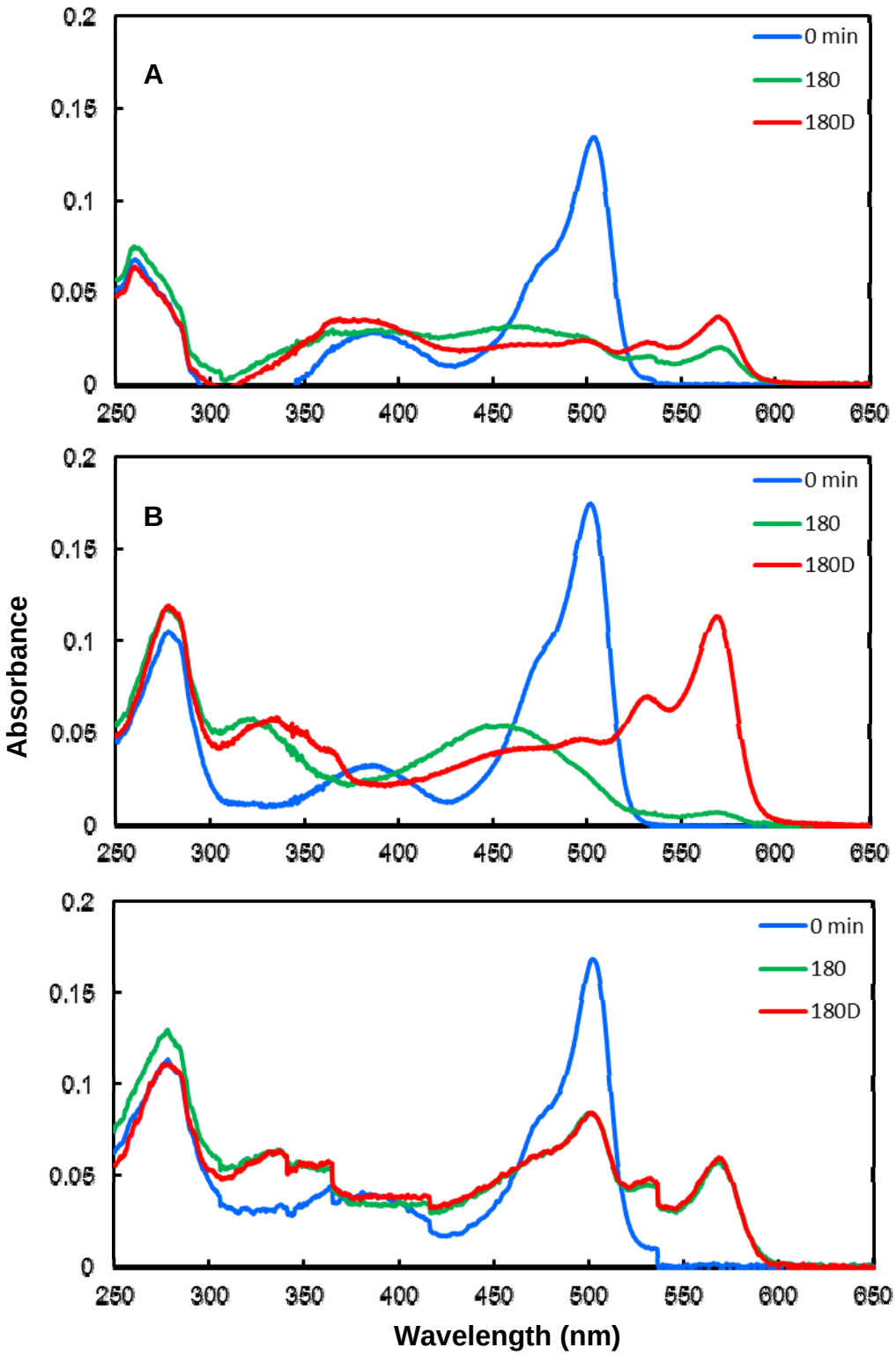


Figure 29. The isomerization of the red LEA chromophore. (A) LEA (B) LEA C204R (C) LEA X(6). (180: after 180 min, 180D: 180 min and overnight in the Dark)

LEA Q38 mutants

In ALL-GFP and LEA variants structures, there is the quadrupolar charge network involving Glu222(211), His206(193), Glu148(144), and Arg69(66) around the chromophore. It turns out that the side chain of Gln42(38) is located above the imidazole ring of His65(62) and makes hydrogen bonds with Glu222(211), and two water molecules (W_a and W_b). Therefore, I hypothesize Gln42(38) could be involved in photoconversion reaction. To evaluate the effect of Q38 on the photoconversion process, three Q38 mutants (Q38A, Q38N, and Q38M) were created and tested. When I introduced Q38 mutations (Q38A, Q38N, and Q38M) into LEA, all three mutants were unable to make the photoconversion reaction (Figure 30). Crystals of all three Q38 mutants were obtained and LEA Q38A structure was solved at 1.95 Å. In LEA Q38A structure, the large-to-small amino acid substitution creates a cavity right beneath His65(62) (Figure 31). Thus, the imidazole ring of His65(62) is rotated towards to the cavity. Also the water molecule W_a is missing, but W_b is located in the same position. The loss of photoconversion ability of Q38 mutants may result from the disruption of a proton transfer chain. However, this structural information cannot completely count out the role of His65 as a proton donor to the carbonyl group of Phe64(61). X-ray crystal structures of the other two Q38 mutants will be a great help to explain the mechanism.

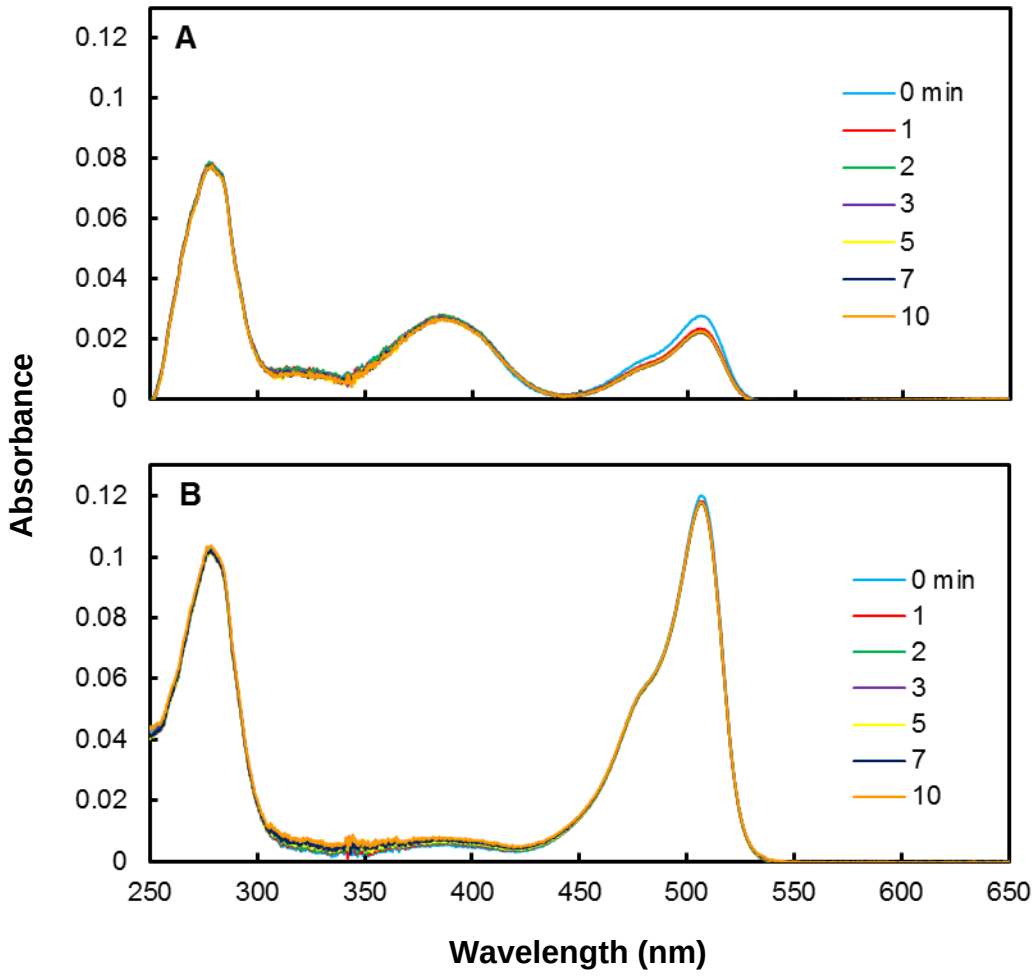


Figure 30. Photoconversion reactions of LEA Q38A. (A) pH 7.01 and (B) pH 9.50 There are little shift between the neutral (380 nm) and the anionic chromophore (500 nm) peaks. However, no red chromophore peaks (480 and 560 nm) are showing.

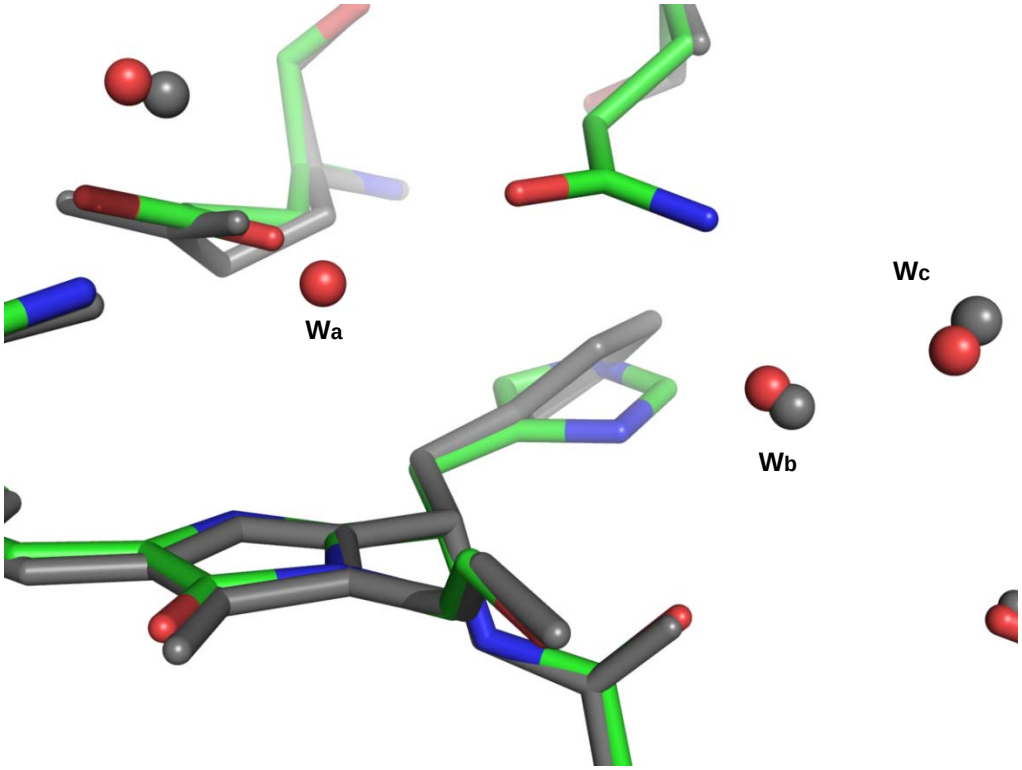


Figure 31. Gln42(38)-Ala mutation changes the conformation of His65(62). In LEA (carbon atoms shown in green), the imidazole ring of His65(62) is planar. In LEA Q38A (all atoms shown in gray), the imidazole ring rotates to the newly created space.

DISCUSSION

In Kaede-type FPs, the conversion from a green-fluorescent (GFP-like) to a red-fluorescent form is activated by UV light. Chromophore excitation facilitates a β-elimination reaction that yields a red-emitting chromophore upon deprotonation of the beta carbon of an active-site histidine residue. The least evolved red ancestor (LEA) is a protein constructed in the laboratory that incorporates the minimum number of 13 substitutions to efficiently generate the red phenotype from the green ancestor, All-GFP. Therefore, we are now in the unique position to examine the structural transitions responsible for efficient color switching. The ancestral reconstruction approach allows us to exclude all ballast mutations, such as those due to evolutionary drift, or due to natural selection for unrelated features. Hence, we can focus on substitutions important for the green-to-red photoconversion. Statistical analysis has demonstrated that 13 substitutions are both necessary and sufficient to obtain the Kaede-type red phenotype from the common ancestor ALL-GFP (*62*). However, any individual reversions from the LEA to the ancestral state, except His65(62), does not completely abolish photoconversion ability. There are four mutations (Q65H, T72A, Y121N, and Y[227-228]del) which are very critical to the photoconversion, whereas the rest of them bring about a relatively small contribution with the large effect residues (A63V, D77H, T109R, S110N, M162T, V165I, R204C, R227H, and M228G) (*62*).

The X-ray crystal structures of ALL-GFP and LEA variants reveal some interesting changes between the non-photoconvertible and photoconvertible proteins. Compared to the non-photoconvertible FPs (ALL-GFP and ALL Q65H),

the photoconvertible FP LEA shows structural differences not only from the inside but also outside of the protein. Two notable changes around the chromophore of LEA are the side chain conformation of Arg69(66) and the additional space resulting from Y121(116)N mutation. On the other hand, the C-terminal chain of LEA is more flexible and does not come in contact with the entrance of the putative solvent channel.

Position of Arg69(66) guanidinium group, LEA X72

The Thr72(69)Ala substitution in LEA allows for direct H-bonding of Arg69(66) to the chromophore carbonyl oxygen, as the reduced side chain size of residue 72 allows for repositioning of the Arg69(66) guanidinium group such that one of the terminal nitrogens interacts with the chromophore, the other is H-bonded to water molecule A (W_a) which in turn interacts with the carboxylate of Glu222(211). Therefore, W_a cannot accept a proton from Glu222(211) to become a hydronium ion, as this would result in electrostatic repulsion from its H-bonding partner Arg69(66). Therefore, proton transfer to W_a is unlikely unless it involves a Groetthius proton-hopping mechanism with concerted transfer to Gln42(38) etc. The reversion variant LEA-X72 that re-introduces a threonine residue likely forces a rearrangement of the Arg69(66) side chain, causing sufficient structural perturbation to cause a rise in chromophore pK_a to 8.4, concomitant with a highly anti-cooperative process that suggests titration of a nearby functional group to its anionic form.

Regardless, hydrogen bonding of Arg69(66) to the chromophore's oxygen is not an essential feature for photoconversion. Recent crystallographic evidence allows for exclusion of Arg69(66) as a critical player in promoting the G/R photochemistry via H-bonding. In Kaede, EosFP and KikGRX, the guanidinium groups of two arginine residues are observed to directly hydrogen bond to the chromophore's carbonyl oxygen, Arg69(66) and Arg96(91) (*42, 50, 51*). Although H-bonding of Arg96(91) appears to be 100% conserved in all fluorescent proteins, the placement of two guanidinium groups in such close proximity to the chromophore's carbonyl oxygen is somewhat unusual, even in fluorescent proteins. However, this type of multiple-arginine interaction is not unique to G/R photoconvertible proteins, as the photoswitchable green fluorescent protein Dronpa exhibits a very similar arrangement without G/R photochemistry (*84*). Most notably, in the green form of Dendra2, Arg69(66) does not hydrogen bond to the chromophore's carbonyl although this variant is highly photoactive in G/R chemistry (*52*). Instead, Arg69(66) forms a salt bridge to Glu211 (in EosFP, Glu222 in GFP), as described previously for several Anthozoa fluorescent proteins (*75, 85, 86*), where this position is occupied by either Arg or Lys. In combination, the crystallographic data clearly repudiate the notion that H-bonding of two guanidinium groups to the chromophore's imidazolinone oxygen is an essential prerequisite for G/R photoconversion.

E1cb mechanism for the photoconversion process

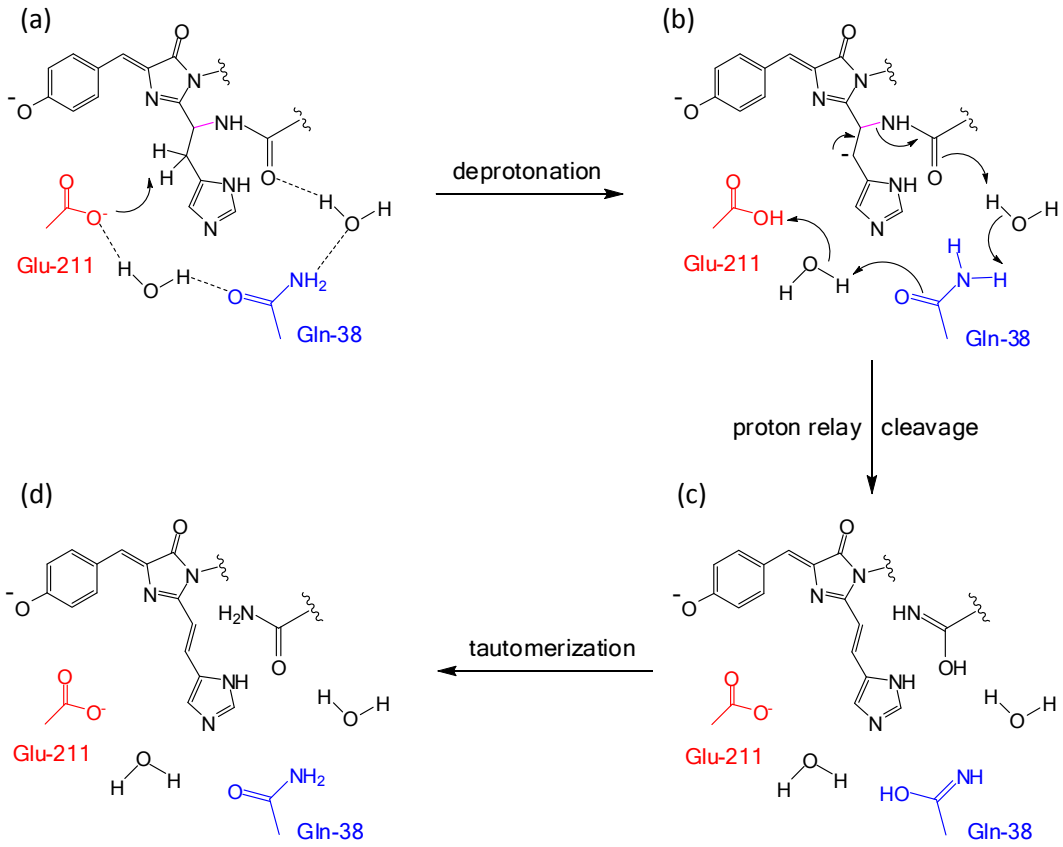
The E1cb mechanism is supported by the proton transfer pathway first identified and described in the present work. In line with previous works, Glu222(211) is

proposed to be the base that initiates the process, generating a carbanion center at C_β of His65 (*42, 87*). The carbanion is highly unstable, and must be neutralized rapidly in order to avoid proton back-transfer from the glutamic acid. Along the same lines of argument, the proton cannot remain on Glu222(211), as immediate back transfer would be favored due to electrostatic disruption of the quadrupolar charge network involving Glu222(211), His206(193), Glu148(144), and Arg69(66). Therefore, rapid proton removal from the carboxylic acid and transfer to some sort of proton sink is an essential prerequisite for a productive mechanism leading to bond scission. In the X-ray structures presented here, as well as all published structures of Kaede-like proteins, no obvious proton sink presents itself, with the possible exception of His65(62). However, the H-bonding network connects with the carbonyl oxygen leaving group, rendering a direct, immediate proton transfer from Glu222(211) to the Phe64(61) backbone oxygen via a proton wire feasible.

Identity of proton donor to the carbonyl oxygen of the leaving group

Protonation of the carboxamide oxygen could occur via imidazolium cation as proposed repeatedly throughout the literature (*50, 51*), or via proton transfer from Glu222(211) via a solvent molecule W_a, the side chain amide of Gln42(38) and another water W_b to the carbonyl oxygen of the leaving group (Scheme 3). This process would require subsequent tautomerization of the Gln42(38) hydroxyimidic acid, a process that would be aided by the water molecule W_b. Here, I propose that the H-bonded network connecting Glu222(211) to Phe64(61) by means of Gln42(38) and two waters (W_a and W_b) is critical in the photoconversion process.

Scheme 3. Proposed mechanism of the green-to-red photoconversion reaction.



In case of amide-iminol tautomerization, the amide form is a predominant tautomer in neutral solutions (K_{eq} = -10^8)(*88*). Considering the photoconversion reaction is not an efficient process, the photoconversion quantum yield for Kaede is 2.4 × 10^{-4}(*41*) and KikGRX is 3.5 × 10^{-3}(*42*), I still propose that a proton transfer via Gln42(38) is a key step in the photoconversion reaction.

A proton transfer involving a tautomerization of Gln has been proposed for a key step for light activation of blue light using flavin adenine dinucleotide (BLUF)-domain proteins (*89-91*). Based on structural data and quantum mechanics calculations, it is proposed that a proton from Tyr21 transfers to the carbonyl oxygen of Gln63 after light absorption and then to N5 of flavin (*91-94*). As a result, Gln63 undergoes tautomerization, resulting in the rearrangement in the hydrogen-bonding network around the flavin chromophore. Although no water molecules are involved in the light activation of BLUF domain proteins, the possibility of involving Gln42(38) provides a good alternative for the answer of the photoconversion reaction.

The fundamental problem with all mechanisms proposed to-date is that they do not address the necessity of rapid proton transfer from Glu222(211) to avoid back-transfer to the His65(62) C_β carbanion. To increase the probability of a productive C_β deprotonation event, a proton sink must exist. This sink could be bulk solvent, some amino acid functional group with appropriate pK_a value, or Phe64(61) carbonyl protonation concerted with C_α-N bond scission. I propose that in LEA, the likelihood of the His65(62) imidazole ring serving as proton sink is somewhat reduced due to the increased distance of W_b to the imidazole ring

nitrogen $N_{\delta 1}$, 4.08 Å in LEA. The separation and geometry does not allow for direct hydrogen bonding and rapid proton transfer (Figure 32). On the other hand, the ALL Q65H crystallographic model shows a distance of 3.24 Å for the same atoms, nearly within H-bonding distance and an angle of about 30° that could provide for weak H-bonding. If the imidazole ring was to serve as a proton sink, the proton would reside on the ring long enough to prevent carbonyl protonation concerted with bond scission. The concerted nature of these two events has been underscored by recent theoretical studies (*87*), and is likely critical to a successful (irreversible) photoconversion process. According to this idea, the position of W_b is intricately linked to photoconversion. The large-to-small mutation Y121(116)N provides the additional space needed for W_b to relax into a position further removed from the imidazole, something that is not possible in the presence of the tyrosine residue, as W_b in Q65H has four H-bonding partners that include Tyr121(116), and is squeezed into a small volume element such that refinement pushes it towards Gln42(38) suggesting a short and strong hydrogen bond, as well as towards His65(62) (Figure 33). The large-to-small substitution involving Y121(116)N is partially compensated for by S105N and A60V, where a slight increase in side chain size allows for improved interior packing interactions.

The role of the histidine residue in photoconversion

Previously proposed hypotheses suggest a proton transfer from the imidazole ring of His65(62) to the carbonyl group of Phe64(61) (*50-52*), Because the mutation on the histidine completely abolishes the photoconversion ability, even with other aromatic amino acids (*37, 39*). However, in our new hypothesis, His65(62)

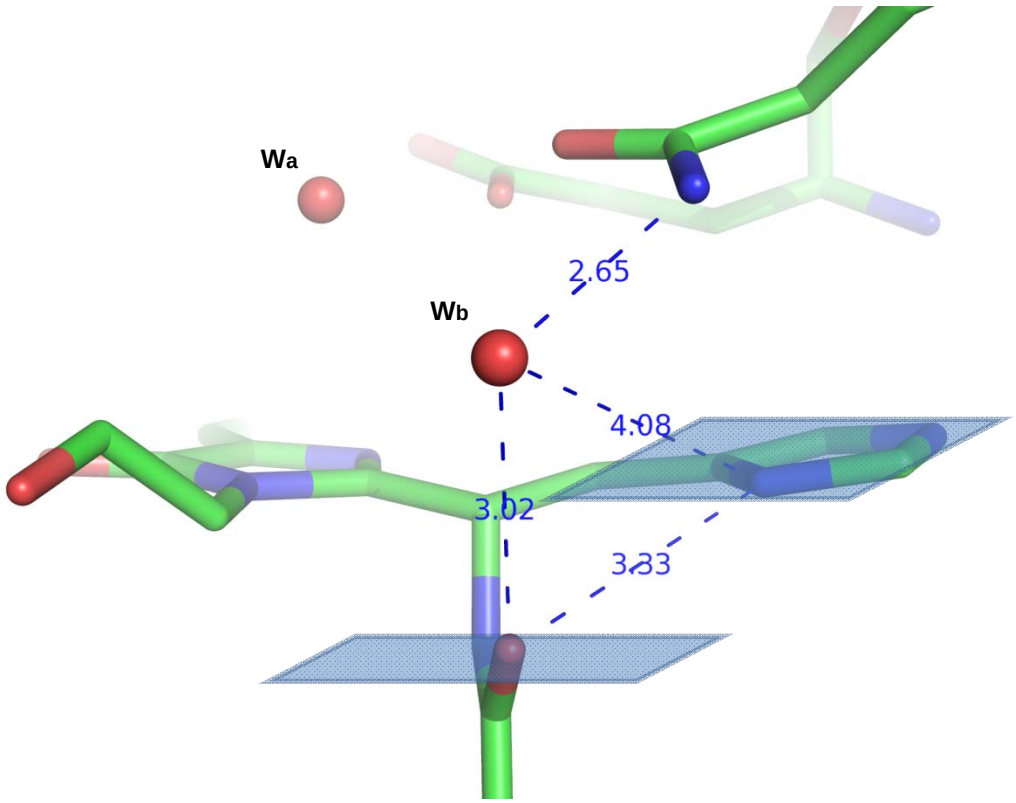
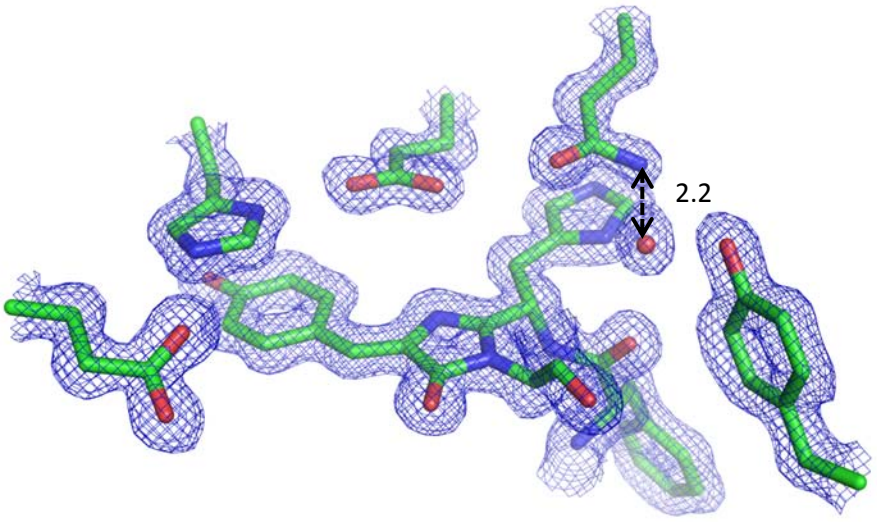


Figure 32. The environment around the carbonyl group of Phe64(61) in the X-ray crystal structure of LEA (distances are shown in angstroms). The plains of the imidazole ring of His65(62) and the peptide bond of Phe64(61) are shown in blue parallelogram.

A



B

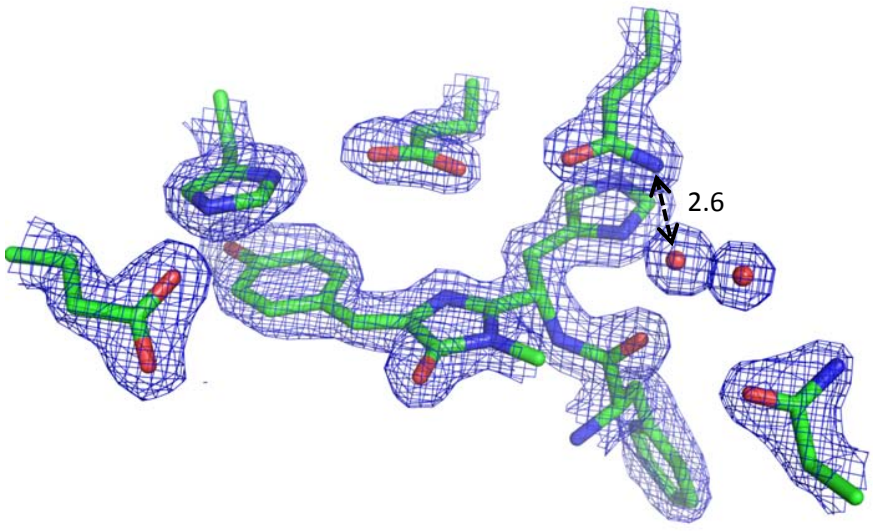


Figure 33. The electron density map ($2F_o$-F_c) of green (A) ALL Q65H and (B) LEA chromophores.

does not donate a proton to the carbonyl group. Here I suggest the role of His65(62) is to provide an imidazole ring for creating an extended π-system of the chromophore. Also it is possible that His assists the tautomerization step by stabilizing a reaction intermediate. The reason why only His can generate photoconversion may be a dense environment around the imidazole ring of chromophore. The pocket around the imidazole ring of His65(62) is very tightly packed (Figure 27), thus any small changes of His65(62) could disrupt a reaction network and affect the photoconversion ability. The last step of photoconversion process is the tautomerization of carboxamide group. Additional water molecule 'W_c' observed crystallographically when Tyr121 is mutated to Asn may facilitate tautomerization of the amide of residue 61 in proton transfer, and tautomerization of the hydroxyimino group to the carboxamide upon leaving group ejection. In LEA Q38A structure, W_a is missing and W_b is still intact (Figure 31). This result shows that W_b interacts strongly with the carbonyl group of Phe64(61) and the additional water molecule 'W_c' can stabilize W_b during the photoconversion reaction.

Is *cis-trans* isomerization necessary for photoconversion?

Photochromic fluorescent proteins undergo reversible photoswitching between a fluorescent (on or light) state and a non-fluorescent (off or dark) state. The structural background of this phenomenon has been studied by crystallographic analysis of several FPs including Dronpa (*80, 84, 95, 96*), Padron (*97*), mTFP0.7 (*31*), asFP595 (*33*), and IrisFP (*43*). In spite of the previous studies, the exact mechanism of photoswitching remains poorly understood. A number of X-ray

crystal structures demonstrated *cis-trans* isomerization of the chromophore (*31, 33, 43, and 95*). However, recent NMR studies showed that off-state switching of Dronpa involves a disordering of the chromophore following light-induced protonation of the phenolic group (*96*).

During the photoconversion experiments, I have observed the photoswitching–like absorbance peak changes between the neutral and anionic chromophores, except LEA X(6). LEA X(6) bears 9 substitution from ALL-GFP, excluding A63(60)V, D77(74)H, M162(154)T, and V165(157)I from 13 LEA mutations (Table 6). In four excluding residues, the residue 165(157) is particularly interesting, which is located in the range of van der Waals contact to the chromophore. On the residue 157, LEA X(6) has Val and LEA has Ile, a larger side chain than Val. Compared to LEA X(6), the small-to-large mutation makes more packed environment around the chromophore of LEA, which could prevent the isomerization (Figure 34). However, LEA X(6) does not show the interchange of peaks. In IrisFP the residue 157 is Ile, the same as LEA. However, to create IrisFP, Phe173-Ser was introduced to EosFP, which is very close to the residue 157. Changes in the chromophore environment induced by Phe173-Ser make the *cis-trans* isomerization possible (*43*). On the other hand, residues 157 and 173 of Dronpa are Val and Phe, respectively. Compared to IrisFP and Dronpa, LEA has bigger amino acids on both residue 157 (Ile) and 173 (Phe). Therefore, there should be another factor to make the isomerization possible in LEA.

The exact nature of *cis-trans* isomerization phenomenon is still unknown. However, this could result from different pKa values of the chromophore between

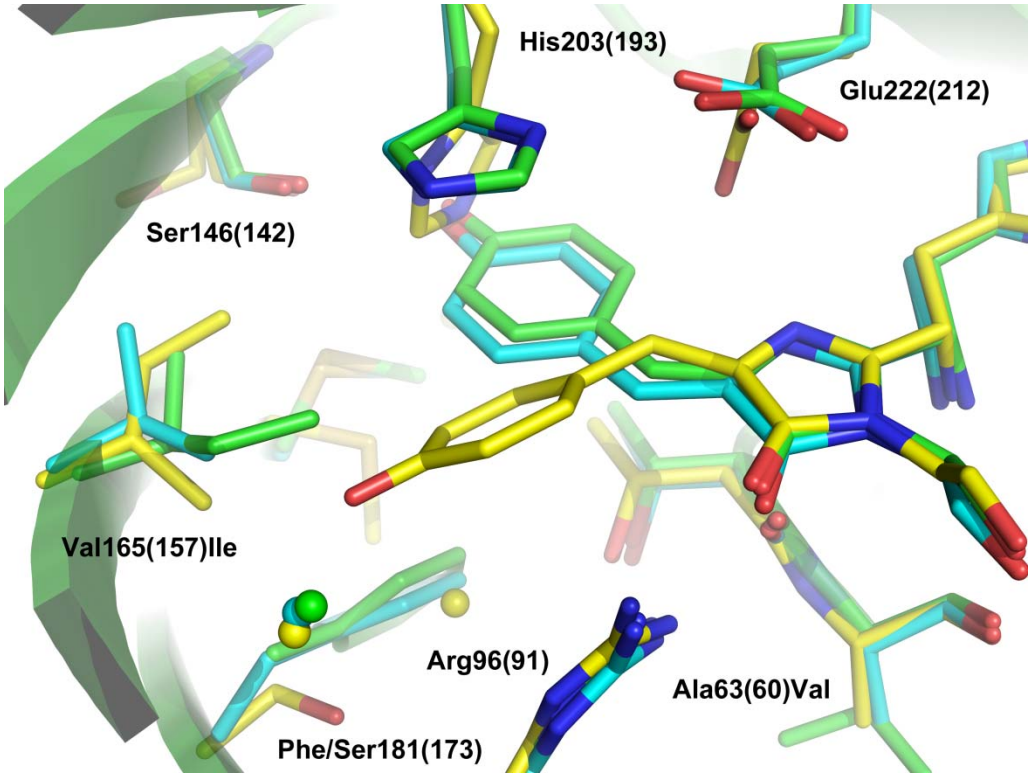


Figure 34. Superposition of the LEA (green), LEA X(6) (cyan), and trans conformation of green IrisFP (yellow).

the ground state and excited state, due to different energy barriers between *cis* and *trans* conformation for protonation-deprotonation equilibrium (*98*). In recent E222Q EGFP study, Bizzarri *et al.* proposed that the isomerization and chromophore ionization are linked to each other. Also they suggested that two types of isomerization process are involved in the photoswitching process, light-induced reversible *cis-trans* isomerization and *trans-cis* only thermal relaxation (*99*).

In the photoconvertible LEA variants, only the isomerization of red chromophore is observed (Figure 29). However, when LEA Q38A was illuminated with UV light, the isomerization of green chromophore was observed (Figure 30). From these results, I speculate that both photoconversion and *cis-trans* isomerization in LEA may be triggered by UV light.

The flexibility of the surface region at the AC dimer

In the photoconversion protein, the lack of C terminus-mediated interaction may lead to changes in flexibility of the solvent channel. This type of difference was previously observed in a structural comparison between 22G and Dronpa (*96*). Dronpa originates from the green fluorescent protein 22G, which includes six substitutions from 22G (Ile102Asn, Phe114Tyr, Leu162Ser, Arg194His, Asn205Ser, and Gly218Glu) (*30*). Compared to the structure of 22G, the on-state Dronpa has a flexible C-terminal chain (*96*). It turns out that two of six substitutions (Arg194His and Gly218Glu) are closely related to the C-terminal chain interaction. In NMR experiments of Dronpa, the loss of the C-terminal chain

interaction makes the chromophore and a part of the β-barrel (mainly β7 and β11 strand) flexible. Interestingly, LEA shares the same 194 (Arg194Cys) mutation and has three more mutations in this area.

Since the ESPT is a first step of the photoconversion process, identifying the path is important to explain the mechanism. The previously proposed ESPT path, from the phenyl ring of Phe66 to the imidazole group of His65, has not been identified from the several X-ray crystal structures of Kaede-type FPs including LEA variants. Therefore, a proton from the phenyl ring of the chromophore must transfer to other acceptors. In other FPs, such proton acceptors are available. In Dronpa, a network of residues, including Ser142, Glu140, His194, Ile195 and a water molecule, may regulate protonation. In IrisFP, Ser142 makes H-bond with the chromophore, as well as a network of water molecules leading into bulk solvent. For LEA, the potential solvent channel may help rapid proton escape after excitation of the chromophore. Proton supply from the solvent is supported by the acid-sensitivity of Dronpa (*30*) and the potential channel also has been found in the on-state Dronpa structure (*80*). As a matter of fact, Kaede-type FPs and Dronpa are members of clade D fluorescent proteins. Therefore, both photoswitching and photoconversion processes may share some common features for the photoactivation.

The enhanced flexibility of the C-terminal chain and the potential solvent channel in the photoconversion FPs are closely linked each other. The flexible β7 and β11 strands could open up the solvent channel and this will potentially improve the ESPT ability in Kaede-type FPs.

The evolution of Kaede-type red fluorescent proteins

In corals, the fluorescent proteins have been evolved from the green fluorescent ancestor (*19, 20*). The color diversification have occurred on several independent events and started from as early as late Triassic – early Jurassic (*16*). The Kaede-type proteins mostly associated with clade D in suborder Faviina, whereas DsRed-type red FPs can be found in all other organisms, such as reef-building corals and all sea anemones (*19*). Most of FPs in suborder Faviina has been evolved from the green ancestor (ALL-GFP). The positions of 13 photoconversion making mutations are located both inside and outside of the β-barrel. After the structural analysis between the reconstructed proteins, I found out a couple of major changes. First, in the photoconvertible FPs, the more flexible C-terminal chain disrupts a direct contact between the entrance of solvent channel and the C-terminal chain itself. This change affects the pK_a values of the chromophore and ESPT. Second, a new space around the carbonyl group of Phe61 was created by Y121(116)N mutation. In the non-photoconvertible FPs, this region is very tightly packed. Therefore, a newly created space with an additional water molecule could provide some room to rearrange during photoconversion reaction. Also other mutations around the chromophore would help to optimize the environment for the process.

It should be noted that not just photoconversion FPs but also photoswitching FPs (dronpa) is found in clade D. At this point, there is no direct evidence that photoconversion and photoswitching mechanisms are related to each other. However, the *cis-trans* isomerization was observed in LEA. Therefore, to identify

common feature between two reactions may give a clue to explain this amazing phenomenon. In addition, it will be interesting to find out what kind of factors exactly make fluorescent proteins in clade D prone to react with light.

Conclusions

Analysis of the molecular evolution of Kaede-type proteins enables us to understand the mechanism of photoconversion process. To answer the questions about evolution of Kaede-type FPs, the ancestral green FP (ALL-GFP), green-to-red photoconvertible FP (LEA), and their variants were reconstructed with ancestral gene reconstruction technique. The structural and spectroscopic study of the ancestral green and photoconversion red FPs provides insights into the structural mechanism and evolution of photoconversion process.

I determined several X-ray crystal structures of the reconstructed proteins. The significant changes for photoconversion process not only occurs the surroundings of the chromophore but also the quaternary structure of protein. First, two 'large-effect' mutations, T72(69)A and Y121(116)N, alter a hydrogen bond network around the chromophore to make photoconversion process possible. Second, the loss of interaction between the C-terminal chain and the solvent channel possibly enhances rapid proton escape from the chromophore. I propose that a new mechanism of photoconversion in Kaede-type FP, which requires a tautomerization of Gln. After the deprotonation of C_β by Glu222(211), the C_β proton is transferred to the carbonyl group of Phe64(61) via Gln42(38). Although the results are not sufficient to completely explain the mechanism of photoconversion, this study contributes to the strategy of rational design of novel fluorescent proteins for efficient molecular probes.

References

1. Remington, S. J. (2002) Negotiating the speed bumps to fluorescence, *Nat Biotechnol 20*, 28-29.

2. Remington, S. J. (2006) Fluorescent proteins: maturation, photochemistry and photophysics, *Curr Opin Struct Biol 16*, 714-721.

3. Wachter, R. M. (2007) Chromogenic cross-link formation in green fluorescent protein, *Acc Chem Res 40*, 120-127.

4. Giepmans, B. N., Adams, S. R., Ellisman, M. H., and Tsien, R. Y. (2006) The fluorescent toolbox for assessing protein location and function, *Science 312*, 217-224.

5. Shaner, N. C., Patterson, G. H., and Davidson, M. W. (2007) Advances in fluorescent protein technology, *J Cell Sci 120*, 4247-4260.

6. Pouwels, L. J., Zhang, L., Chan, N. H., Dorrestein, P. C., and Wachter, R. M. (2008) Kinetic isotope effect studies on the de novo rate of chromophore formation in fast- and slow-maturing GFP variants, *Biochemistry 47*, 10111-10122.

7. Wachter, R. M., Watkins, J. L., and Kim, H. (2010) Mechanistic diversity of red fluorescence acquisition by GFP-like proteins, *Biochemistry 49*, 7417-7427.

8. Sniegowski, J. A., Lappe, J. W., Patel, H. N., Huffman, H. A., and Wachter, R. M. (2005) Base catalysis of chromophore formation in Arg96 and Glu222 variants of green fluorescent protein, *J Biol Chem 280*, 26248-26255.

9. Sniegowski, J. A., Phail, M. E., and Wachter, R. M. (2005) Maturation efficiency, trypsin sensitivity, and optical properties of Arg96, Glu222, and Gly67 variants of green fluorescent protein, *Biochem Biophys Res Commun 332*, 657-663.

10. Zhang, L., Patel, H. N., Lappe, J. W., and Wachter, R. M. (2006) Reaction progress of chromophore biogenesis in green fluorescent protein, *J Am Chem Soc 128*, 4766-4772.

11. Rosenow, M. A., Huffman, H. A., Phail, M. E., and Wachter, R. M. (2004) The crystal structure of the Y66L variant of green fluorescent protein supports a cyclization-oxidation-dehydration mechanism for chromophore maturation, *Biochemistry 43*, 4464-4472.

12. Rosenow, M. A., Patel, H. N., and Wachter, R. M. (2005) Oxidative chemistry in the GFP active site leads to covalent cross-linking of a modified leucine side chain with a histidine imidazole: implications for the mechanism of chromophore formation, *Biochemistry 44*, 8303-8311.

13. Hopf, M., Gohring, W., Ries, A., Timpl, R., and Hohenester, E. (2001) Crystal structure and mutational analysis of a perlecan-binding fragment of nidogen-1, *Nat Struct Biol 8*, 634-640.

14. Matz, M. V., Labas, Y. A., and Ugalde, J. (2006) Evolution of function and color in GFP-like proteins, *Methods Biochem Anal 47*, 139-161.

15. Matz, M. V., Marshall, N. J., and Vorobyev, M. (2006) Are corals colorful?, *Photochem Photobiol 82*, 345-350.

16. Kelmanson, I. V., and Matz, M. V. (2003) Molecular basis and evolutionary origins of color diversity in great star coral Montastraea cavernosa (Scleractinia: Faviida), *Mol Biol Evol 20*, 1125-1133.

17. Oswald, F., Schmitt, F., Leutenegger, A., Ivanchenko, S., D'Angelo, C., Salih, A., Maslakova, S., Bulina, M., Schirmbeck, R., Nienhaus, G. U., Matz, M. V., and Wiedenmann, J. (2007) Contributions of host and symbiont pigments to the coloration of reef corals, *FEBS J 274*, 1102-1109.

18. Labas, Y. A., Gurskaya, N. G., Yanushevich, Y. G., Fradkov, A. F., Lukyanov, K. A., Lukyanov, S. A., and Matz, M. V. (2002) Diversity and evolution of the green fluorescent protein family, *Proc Natl Acad Sci U S A 99*, 4256-4261.

19. Alieva, N. O., Konzen, K. A., Field, S. F., Meleshkevitch, E. A., Hunt, M. E., Beltran-Ramirez, V., Miller, D. J., Wiedenmann, J., Salih, A., and Matz, M. V. (2008) Diversity and evolution of coral fluorescent proteins, *PLoS One 3*, e2680.

20. Ugalde, J. A., Chang, B. S., and Matz, M. V. (2004) Evolution of coral pigments recreated, *Science 305*, 1433.

21. Field, S. F., Bulina, M. Y., Kelmanson, I. V., Bielawski, J. P., and Matz, M. V. (2006) Adaptive evolution of multicolored fluorescent proteins in reef-building corals, *J Mol Evol 62*, 332-339.

22. Lin, M. Z., McKeown, M. R., Ng, H. L., Aguilera, T. A., Shaner, N. C., Campbell, R. E., Adams, S. R., Gross, L. A., Ma, W., Alber, T., and Tsien, R. Y. (2009) Autofluorescent proteins with excitation in the optical window for intravital imaging in mammals, *Chem Biol 16*, 1169-1179.

23. Ebbinghaus, S., Dhar, A., McDonald, J. D., and Gruebele, M. (2010) Protein folding stability and dynamics imaged in a living cell, *Nat Methods 7*, 319-323.

24. Piatkevich, K. D., and Verkhusha, V. V. (2010) Advances in engineering of fluorescent proteins and photoactivatable proteins with red emission, *Curr Opin Chem Biol 14*, 23-29.

25. Lukyanov, K. A., Chudakov, D. M., Lukyanov, S., and Verkhusha, V. V. (2005) Innovation: Photoactivatable fluorescent proteins, *Nat Rev Mol Cell Biol 6*, 885-891.

26. Wiedenmann, J., and Nienhaus, G. U. (2006) Live-cell imaging with EosFP and other photoactivatable marker proteins of the GFP family, *Expert Rev Proteomics 3*, 361-374.

27. Hess, S. T., Girirajan, T. P., and Mason, M. D. (2006) Ultra-high resolution imaging by fluorescence photoactivation localization microscopy, *Biophys J 91*, 4258-4272.

28. Betzig, E., Patterson, G. H., Sougrat, R., Lindwasser, O. W., Olenych, S., Bonifacino, J. S., Davidson, M. W., Lippincott-Schwartz, J., and Hess, H. F. (2006) Imaging intracellular fluorescent proteins at nanometer resolution, *Science 313*, 1642-1645.

29. Hofmann, M., Eggeling, C., Jakobs, S., and Hell, S. W. (2005) Breaking the diffraction barrier in fluorescence microscopy at low light intensities by using reversibly photoswitchable proteins, *Proc Natl Acad Sci U S A 102*, 17565-17569.

30. Ando, R., Mizuno, H., and Miyawaki, A. (2004) Regulated fast nucleocytoplasmic shuttling observed by reversible protein highlighting, *Science 306*, 1370-1373.

31. Henderson, J. N., Ai, H. W., Campbell, R. E., and Remington, S. J. (2007) Structural basis for reversible photobleaching of a green fluorescent protein homologue, *Proc Natl Acad Sci U S A 104*, 6672-6677.

32. Lukyanov, K. A., Fradkov, A. F., Gurskaya, N. G., Matz, M. V., Labas, Y. A., Savitsky, A. P., Markelov, M. L., Zaraisky, A. G., Zhao, X., Fang, Y., Tan, W., and Lukyanov, S. A. (2000) Natural animal coloration can Be determined by a nonfluorescent green fluorescent protein homolog, *J Biol Chem 275*, 25879-25882.

33. Andresen, M., Wahl, M. C., Stiel, A. C., Grater, F., Schafer, L. V., Trowitzsch, S., Weber, G., Eggeling, C., Grubmuller, H., Hell, S. W., and Jakobs, S. (2005) Structure and mechanism of the reversible photoswitch of a fluorescent protein, *Proc Natl Acad Sci U S A 102*, 13070-13074.

34. Patterson, G. H., and Lippincott-Schwartz, J. (2002) A photoactivatable GFP for selective photolabeling of proteins and cells, *Science 297*, 1873-1877.

35. Verkhusha, V. V., and Sorkin, A. (2005) Conversion of the monomeric red fluorescent protein into a photoactivatable probe, *Chem Biol 12*, 279-285.

36. Chudakov, D. M., Verkhusha, V. V., Staroverov, D. B., Souslova, E. A., Lukyanov, S., and Lukyanov, K. A. (2004) Photoswitchable cyan fluorescent protein for protein tracking, *Nat Biotechnol 22*, 1435-1439.

37. Mizuno, H., Mal, T. K., Tong, K. I., Ando, R., Furuta, T., Ikura, M., and Miyawaki, A. (2003) Photo-induced peptide cleavage in the green-to-red conversion of a fluorescent protein, *Mol Cell 12*, 1051-1058.

38. Tsutsui, H., Karasawa, S., Shimizu, H., Nukina, N., and Miyawaki, A. (2005) Semi-rational engineering of a coral fluorescent protein into an efficient highlighter, *EMBO Rep 6*, 233-238.

39. Wiedenmann, J., Ivanchenko, S., Oswald, F., Schmitt, F., Rocker, C., Salih, A., Spindler, K. D., and Nienhaus, G. U. (2004) EosFP, a fluorescent marker protein with UV-inducible green-to-red fluorescence conversion, *Proc Natl Acad Sci U S A 101*, 15905-15910.

40. Gurskaya, N. G., Verkhusha, V. V., Shcheglov, A. S., Staroverov, D. B., Chepurnykh, T. V., Fradkov, A. F., Lukyanov, S., and Lukyanov, K. A. (2006) Engineering of a monomeric green-to-red photoactivatable fluorescent protein induced by blue light, *Nat Biotechnol 24*, 461-465.

41. Ando, R., Hama, H., Yamamoto-Hino, M., Mizuno, H., and Miyawaki, A. (2002) An optical marker based on the UV-induced green-to-red photoconversion of a fluorescent protein, *Proc Natl Acad Sci U S A 99*, 12651-12656.

42. Tsutsui, H., Shimizu, H., Mizuno, H., Nukina, N., Furuta, T., and Miyawaki, A. (2009) The E1 mechanism in photo-induced beta-elimination reactions for green-to-red conversion of fluorescent proteins, *Chem Biol 16*, 1140-1147.

43. Adam, V., Lelimousin, M., Boehme, S., Desfonds, G., Nienhaus, K., Field, M. J., Wiedenmann, J., McSweeney, S., Nienhaus, G. U., and Bourgeois, D. (2008) Structural characterization of IrisFP, an optical highlighter undergoing multiple photo-induced transformations, *Proc Natl Acad Sci U S A 105*, 18343-18348.

44. Pakhomov, A. A., Martynova, N. Y., Gurskaya, N. G., Balashova, T. A., and Martynov, V. I. (2004) Photoconversion of the chromophore of a fluorescent protein from Dendronephthya sp, *Biochemistry (Mosc) 69*, 901-908.

45. Tsien, R. Y. (1998) The green fluorescent protein, *Annu Rev Biochem 67*, 509-544.

46. Elsliger, M. A., Wachter, R. M., Hanson, G. T., Kallio, K., and Remington, S. J. (1999) Structural and spectral response of green fluorescent protein variants to changes in pH, *Biochemistry 38*, 5296-5301.

47. Wachter, R. M., Yarbrough, D., Kallio, K., and Remington, S. J. (2000) Crystallographic and energetic analysis of binding of selected anions to the yellow variants of green fluorescent protein, *J Mol Biol 301*, 157-171.

48. Chattoraj, M., King, B. A., Bublitz, G. U., and Boxer, S. G. (1996) Ultra-fast excited state dynamics in green fluorescent protein: multiple states and proton transfer, *Proc Natl Acad Sci U S A 93*, 8362-8367.

49. Hosoi, H., Mizuno, H., Miyawaki, A., and Tahara, T. (2006) Competition between energy and proton transfer in ultrafast excited-state dynamics of an oligomeric fluorescent protein red Kaede, *J Phys Chem B 110*, 22853-22860.

50. Nienhaus, K., Nienhaus, G. U., Wiedenmann, J., and Nar, H. (2005) Structural basis for photo-induced protein cleavage and green-to-red conversion of fluorescent protein EosFP, *Proc Natl Acad Sci U S A 102*, 9156-9159.

51. Hayashi, I., Mizuno, H., Tong, K. I., Furuta, T., Tanaka, F., Yoshimura, M., Miyawaki, A., and Ikura, M. (2007) Crystallographic evidence for water-assisted photo-induced peptide cleavage in the stony coral fluorescent protein Kaede, *J Mol Biol 372*, 918-926.

52. Adam, V., Nienhaus, K., Bourgeois, D., and Nienhaus, G. U. (2009) Structural basis of enhanced photoconversion yield in green fluorescent protein-like protein Dendra2, *Biochemistry 48*, 4905-4915.

53. Lelimousin, M., Adam, V., Nienhaus, G. U., Bourgeois, D., and Field, M. J. (2009) Photoconversion of the fluorescent protein EosFP: a hybrid potential simulation study reveals intersystem crossings, *J Am Chem Soc 131*, 16814-16823.

54. Thornton, J. W. (2004) Resurrecting ancient genes: experimental analysis of extinct molecules, *Nat Rev Genet 5*, 366-375.

55. Harms, M. J., and Thornton, J. W. (2010) Analyzing protein structure and function using ancestral gene reconstruction, *Curr Opin Struct Biol 20*, 360-366.

56. Peisajovich, S. G., and Tawfik, D. S. (2007) Protein engineers turned evolutionists, *Nat Methods 4*, 991-994.

57. Chandrasekharan, U. M., Sanker, S., Glynias, M. J., Karnik, S. S., and Husain, A. (1996) Angiotensin II-forming activity in a reconstructed ancestral chymase, *Science 271*, 502-505.

58. Zhang, J., and Rosenberg, H. F. (2002) Complementary advantageous substitutions in the evolution of an antiviral RNase of higher primates, *Proc Natl Acad Sci U S A 99*, 5486-5491.

59. Bridgham, J. T., Carroll, S. M., and Thornton, J. W. (2006) Evolution of hormone-receptor complexity by molecular exploitation, *Science 312*, 97-101.

60. Ortlund, E. A., Bridgham, J. T., Redinbo, M. R., and Thornton, J. W. (2007) Crystal structure of an ancient protein: evolution by conformational epistasis, *Science 317*, 1544-1548.

61. Bridgham, J. T., Ortlund, E. A., and Thornton, J. W. (2009) An epistatic ratchet constrains the direction of glucocorticoid receptor evolution, *Nature 461*, 515-519.

62. Field, S. F., and Matz, M. V. (2010) Retracing evolution of red fluorescence in GFP-like proteins from Faviina corals, *Mol Biol Evol 27*, 225-233.

63. Chang, B. S., Ugalde, J. A., and Matz, M. V. (2005) Applications of ancestral protein reconstruction in understanding protein function: GFP-like proteins, *Methods Enzymol 395*, 652-670.

64. Shagin, D. A., Barsova, E. V., Yanushevich, Y. G., Fradkov, A. F., Lukyanov, K. A., Labas, Y. A., Semenova, T. N., Ugalde, J. A., Meyers, A., Nunez, J. M., Widder, E. A., Lukyanov, S. A., and Matz, M. V. (2004) GFP-like proteins as ubiquitous metazoan superfamily: evolution of functional features and structural complexity, *Mol Biol Evol 21*, 841-850.

65. Pei, J., Tang, M., and Grishin, N. V. (2008) PROMALS3D web server for accurate multiple protein sequence and structure alignments, *Nucleic Acids Res 36*, W30-34.

66. Leslie, A. G. (1999) Integration of macromolecular diffraction data, *Acta Crystallogr D Biol Crystallogr 55*, 1696-1702.

67. (1994) The CCP4 suite: programs for protein crystallography, *Acta Crystallogr D Biol Crystallogr 50*, 760-763.

68. McCoy, A. J. (2007) Solving structures of protein complexes by molecular replacement with Phaser, *Acta Crystallogr D Biol Crystallogr 63*, 32-41.

69. Zwart, P. H., Langer, G. G., and Lamzin, V. S. (2004) Modelling bound ligands in protein crystal structures, *Acta Crystallogr D Biol Crystallogr 60*, 2230-2239.

70. Emsley, P., and Cowtan, K. (2004) Coot: model-building tools for molecular graphics, *Acta Crystallogr D Biol Crystallogr 60*, 2126-2132.

71. Murshudov, G. N., Vagin, A. A., and Dodson, E. J. (1997) Refinement of macromolecular structures by the maximum-likelihood method, *Acta Crystallogr D Biol Crystallogr 53*, 240-255.

72. Nienhaus, K., Renzi, F., Vallone, B., Wiedenmann, J., and Nienhaus, G. U. (2006) Exploring chromophore--protein interactions in fluorescent protein cmFP512 from Cerianthus membranaceus: X-ray structure analysis and optical spectroscopy, *Biochemistry 45*, 12942-12953.

73. Otwinowski, Z. M., W. (1997) *Processing of X-ray diffraction data collected in oscillation mode*, Vol. 276, Academic Press, New York.

74. Segel, I. H. (1975) *Enzyme kinetics : behavior and analysis of rapid equilibrium and steady state enzyme systems / Irwin H. Segel.*, Wiley, New York.

75. Yarbrough, D., Wachter, R. M., Kallio, K., Matz, M. V., and Remington, S. J. (2001) Refined crystal structure of DsRed, a red fluorescent protein from coral, at 2.0-A resolution, *Proc Natl Acad Sci U S A 98*, 462-467.

76. Krissinel, E., and Henrick, K. (2004) Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions, *Acta Crystallogr D Biol Crystallogr 60*, 2256-2268.

77. Evdokimov, A. G., Pokross, M. E., Egorov, N. S., Zaraisky, A. G., Yampolsky, I. V., Merzlyak, E. M., Shkoporov, A. N., Sander, I., Lukyanov, K. A., and Chudakov, D. M. (2006) Structural basis for the fast maturation of Arthropoda green fluorescent protein, *EMBO Rep 7*, 1006-1012.

78. Pletneva, N., Pletnev, S., Tikhonova, T., Popov, V., Martynov, V., and Pletnev, V. (2006) Structure of a red fluorescent protein from Zoanthus, zRFP574, reveals a novel chromophore, *Acta Crystallogr D Biol Crystallogr 62*, 527-532.

79. Pletnev, S., Shcherbo, D., Chudakov, D. M., Pletneva, N., Merzlyak, E. M., Wlodawer, A., Dauter, Z., and Pletnev, V. (2008) A crystallographic study of bright far-red fluorescent protein mKate reveals pH-induced cis-trans isomerization of the chromophore, *J Biol Chem 283*, 28980-28987.

80. Stiel, A. C., Trowitzsch, S., Weber, G., Andresen, M., Eggeling, C., Hell, S. W., Jakobs, S., and Wahl, M. C. (2007) 1.8 A bright-state structure of the reversibly switchable fluorescent protein Dronpa guides the generation of fast switching variants, *Biochem J 402*, 35-42.

81. Petrek, M., Otyepka, M., Banas, P., Kosinova, P., Koca, J., and Damborsky, J. (2006) CAVER: a new tool to explore routes from protein clefts, pockets and cavities, *BMC Bioinformatics 7*, 316.

82. Andresen, M., Stiel, A. C., Folling, J., Wenzel, D., Schonle, A., Egner, A., Eggeling, C., Hell, S. W., and Jakobs, S. (2008) Photoswitchable fluorescent proteins enable monochromatic multilabel imaging and dual color fluorescence nanoscopy, *Nat Biotechnol 26*, 1035-1040.

83. Voliani, V., Bizzarri, R., Nifosi, R., Abbruzzetti, S., Grandi, E., Viappiani, C., and Beltram, F. (2008) Cis-trans photoisomerization of fluorescent-protein chromophores, *J Phys Chem B 112*, 10714-10722.

84. Wilmann, P. G., Turcic, K., Battad, J. M., Wilce, M. C., Devenish, R. J., Prescott, M., and Rossjohn, J. (2006) The 1.7 A crystal structure of Dronpa: a photoswitchable green fluorescent protein, *J Mol Biol 364*, 213-224.

85. Henderson, J. N., and Remington, S. J. (2005) Crystal structures and mutational analysis of amFP486, a cyan fluorescent protein from Anemonia majano, *Proc Natl Acad Sci U S A 102*, 12712-12717.

86. Malo, G. D., Wang, M., Wu, D., Stelling, A. L., Tonge, P. J., and Wachter, R. M. (2008) Crystal structure and Raman studies of dsFP483, a cyan fluorescent protein from Discosoma striata, *J Mol Biol 378*, 871-886.

87. Li, X., Chung, L. W., Mizuno, H., Miyawaki, A., and Morokuma, K. (2010) Competitive mechanistic pathways for green-to-red photoconversion in the fluorescent protein Kaede: a computational study, *J Phys Chem B 114*, 16666-16675.

88. Sigel, H., and Martin, R. B. (1982) Coordinating Properties of the Amide Bond. Stability and Structure of Metal Ion Complexes of Peptides and Related Ligands, *Chemical Reviews 82*, 385-426.

89. Masuda, S., and Bauer, C. E. (2002) AppA is a blue light photoreceptor that antirepresses photosynthesis gene expression in Rhodobacter sphaeroides, *Cell 110*, 613-623.

90. Laan, W., Gauden, M., Yeremenko, S., van Grondelle, R., Kennis, J. T., and Hellingwerf, K. J. (2006) On the mechanism of activation of the BLUF domain of AppA, *Biochemistry 45*, 51-60.

91. Stelling, A. L., Ronayne, K. L., Nappa, J., Tonge, P. J., and Meech, S. R. (2007) Ultrafast structural dynamics in BLUF domains: transient infrared spectroscopy of AppA and its mutants, *J Am Chem Soc 129*, 15556-15564.

92. Domratcheva, T., Grigorenko, B. L., Schlichting, I., and Nemukhin, A. V. (2008) Molecular models predict light-induced glutamine tautomerization in BLUF photoreceptors, *Biophys J 94*, 3872-3879.

93. Sadeghian, K., Bocola, M., and Schutz, M. (2008) A conclusive mechanism of the photoinduced reaction cascade in blue light using flavin photoreceptors, *J Am Chem Soc 130*, 12501-12513.

94. Sadeghian, K., Bocola, M., and Schutz, M. (2010) A QM/MM study on the fast photocycle of blue light using flavin photoreceptors in their light-adapted/active form, *Phys Chem Chem Phys 12*, 8840-8846.

95. Andresen, M., Stiel, A. C., Trowitzsch, S., Weber, G., Eggeling, C., Wahl, M. C., Hell, S. W., and Jakobs, S. (2007) Structural basis for reversible photoswitching in Dronpa, *Proc Natl Acad Sci U S A 104*, 13005-13009.

96. Mizuno, H., Mal, T. K., Walchli, M., Kikuchi, A., Fukano, T., Ando, R., Jeyakanthan, J., Taka, J., Shiro, Y., Ikura, M., and Miyawaki, A. (2008) Light-dependent regulation of structural flexibility in a photochromic fluorescent protein, *Proc Natl Acad Sci U S A 105*, 9227-9232.

97. Brakemann, T., Weber, G., Andresen, M., Groenhof, G., Stiel, A. C., Trowitzsch, S., Eggeling, C., Grubmuller, H., Hell, S. W., Wahl, M. C., and Jakobs, S. (2010) Molecular basis of the light-driven switching of the photochromic fluorescent protein Padron, *J Biol Chem 285*, 14603-14609.

98. He, X., Bell, A. F., and Tonge, P. J. (2003) Ground state isomerization of a model green fluorescent protein chromophore, *FEBS Lett 549*, 35-38.

99. Bizzarri, R., Serresi, M., Cardarelli, F., Abbruzzetti, S., Campanini, B., Viappiani, C., and Beltram, F. (2010) Single amino acid replacement makes Aequorea victoria fluorescent proteins reversibly photoswitchable, *J Am Chem Soc 132*, 85-95.